

HETEROGENEOUS NETWORK EPIDEMICS: REAL-TIME GROWTH, VARIANCE AND EXTINCTION OF INFECTION

Frank Ball

Thomas House

Abstract

Recent years have seen a large amount of interest in epidemics on networks as a way of representing the complex structure of contacts capable of spreading infections through the modern human population. The configuration model is a popular choice in theoretical studies since it combines the ability to specify the distribution of the number of contacts (degree) with analytical tractability. Here we consider the early real-time behaviour of the Markovian SIR epidemic model on a configuration model network using a multi-type branching process. We find closed-form analytic expressions for the mean and variance of the number of infectious individuals as a function of time and the degree of the initially infected individual(s), and write down a system of differential equations for the probability of extinction that are numerically fast compared to Monte Carlo simulation. We show that these quantities are all sensitive to the degree distribution – in particular we confirm that the mean prevalence of infection depends on the first two moments of the degree distribution and the variance in prevalence depends on the first three moments of the degree distribution. In contrast to most existing analytic approaches, the accuracy of these results does not depend on having a large number of infectious individuals, meaning that in the large population limit they would be asymptotically exact even for one initial infectious individual.

Keywords: SIR epidemic; Configuration model; Branching process

1 Introduction

1.1 Background

Models of infectious disease transmission have, from relatively modest beginnings (e.g. Bailey [2]), developed a rich domain of applicability covering the whole spectrum of human, animal and plant pathogens, and informing the study of questions from viral evolution, through epidemiology of infectious diseases, to public health policy (see Heesterbeek *et al.* [17]). Increasingly, networks have been seen as a way of modelling the complex, heterogeneous patterns of contacts between individuals (Danon *et al.* [10]).

In theoretical studies, the configuration model has been a popular choice due to the ability to specify the number of contacts each individual has that are capable of spreading disease, while allowing for analytic results to be obtained (e.g. Molloy and Reed [25] and Newman [27]).

In a recent paper, Graham and House [16] use a pairwise approximation in conjunction with the central limit theorem for density dependent population processes (Ethier and Kurtz [15], Chapter 11) to obtain a closed-form approximation to the mean and variance of prevalence in the linearised model which approximates the early asymptotic exponential growth phase of a Markovian SIR epidemic on a configuration network. In particular, they find that, under these approximations, the variance in disease prevalence is determined

by the first three moments of the network degree distribution. In this paper, we use the effective degree approach of Ball and Neal [4] to approximate the early stages of the epidemic by a continuous-time, multitype Markovian branching process, which is then analysed in detail. For $t > 0$, let $Z(t)$ denote the total number of individuals alive in this branching process at time t , so $Z(t)$ approximates disease prevalence in the epidemic model during its early asymptotic exponential growth phase. Explicit closed-form expressions are derived for the mean and variance of $Z(t)$, the covariance of $Z(t)$ and $Z(s)$ to give the behaviour over time, and also for the probability of extinction $\pi(t) = \mathbb{P}(Z(t) = 0)$. As in Graham and House [16], the mean and variance in disease prevalence depends on the degree distribution only through its first two and three moments, respectively.

The results in Graham and House [16] assume implicitly that the initial number of infectives is sufficiently large for the density dependent population process central limit theorem to yield a good approximation. In contrast, our results assume any arbitrary, but specified, initial number of infectives. The asymptotic distribution of types in the branching process, when it does not go extinct, is also available in closed-form and enables us to obtain a Gaussian process approximation, with explicit mean and covariance function, for the prevalence in the early asymptotic exponential growth phase of an SIR epidemic, with few initial infectives, which takes off and becomes established. We show that this approximation can be applied together with the methods of Ross *et al.* [30] to estimate epidemiological parameters from early prevalence data of a simulated epidemic provided the degree distribution is known.

1.2 Outline of the paper

The paper is organised as follows. The configuration network model and a Markov SIR epidemic on that network are described in Section 2.1. The effective degree construction of this epidemic is outlined in Section 2.2 and the continuous-time, multitype Markov branching process which approximates the early stages of the epidemic is introduced in Section 2.3. The mean, variance and covariance functions of the total number of individuals alive in the branching process are considered in Sections 3, 4 and 5, respectively. Explicit closed-form expressions are obtained for each of these quantities and for their limits as time $t \rightarrow \infty$. The arguments in Sections 3, 4 and 5 assume that underlying degree distribution has a maximum degree. In Section 6, we show that these expressions continue to hold in the unbounded degree setting, subject to the degree distribution satisfying suitable moment conditions. The probability that the branching process is extinct at time t is studied in Section 7. Closed-form expressions for this probability, given the initial state of the branching process, are not available so asymptotic results as $t \rightarrow 0$ and $t \rightarrow \infty$ are considered.

The mean, variance and covariance functions derived in Sections 3, 4 and 5 are unconditional, so they include realisations of the branching process which result in extinction. However, in the epidemic setting, we are often interested in analysing the behaviour of epidemics that take off and become established, which correspond to non-extinction of the branching process. In Section 8, we first derive the mean and variance of the total number of individuals alive in the branching process at time t , conditional upon the process having survived to time t ; fully closed-form results are not available owing to the absence of a closed-form expression for the survival probability. We then consider realisations of the branching process which reach some specified size, K say, with time being set to zero the first time the total number of individuals alive is K . The results in Section 3 yield an explicit expression for the asymptotic distribution of types, given that the branching

process does not go extinct, which, provided K is sufficiently large, enables the above branching process starting from K individuals to be approximated by a Gaussian process whose mean and covariance functions are determined explicitly. The theory is illustrated by numerical examples of both forward simulation and inference in Section 9 and some concluding comments are given in Section 10.

2 Model and approximating branching process

2.1 Model

We consider the spread of an SIR epidemic on a network of N individuals, labelled $1, 2, \dots, N$, constructed using the configuration model as follows (see e.g. Newman [27]). Let D be a random variable which describes the degree of a typical individual and let $d_k = \mathbb{P}(D = k)$ ($k = 0, 1, \dots$). Let D_1, D_2, \dots, D_N be independent realisations of D and, for $i = 1, 2, \dots, N$, attach D_i stubs (half-edges) to individual i . Pair up these stubs uniformly at random to form the edges in the network. If $D_1 + D_2 + \dots + D_N$ is odd, there will be a left-over stub, which is ignored; the resulting network may have other ‘defects’ such as self-loops and multiple edges between pairs of individuals but, provided that D has finite variance, such imperfections become sparse in the network as $N \rightarrow \infty$ (see e.g. Durrett [13], Theorem 3.1.2). An alternative to the degrees D_1, D_2, \dots, D_N being random (see e.g. Molloy and Reed [25]) is, for each $N = 1, 2, \dots$, to replace $\mathbf{D} = (D_1, D_2, \dots, D_N)$ by $\mathbf{D}^{(N)} = (D_1^{(N)}, D_2^{(N)}, \dots, D_N^{(N)})$, where the degree sequences $\mathbf{D}^{(N)}$ ($N = 1, 2, \dots$) are prescribed and satisfy $N^{-1} \sum_{i=1}^N \delta_{k, D_i^{(N)}} \rightarrow d_k$ as $n \rightarrow \infty$ ($k = 0, 1, \dots$), where the Kronecker delta $\delta_{k,j}$ is 1 if $k = j$ and 0 otherwise. The results of the paper all apply also to this setting.

The epidemic is defined as follows. Initially some individuals are infective and the remaining individuals are susceptible. Infective individuals have independent infectious periods, each having an exponential distribution with rate γ (and hence mean γ^{-1}), after which they become recovered and play no further role in the epidemic. Throughout its infectious period, an infective contacts each of its susceptible neighbours in the network at the points of independent Poisson processes having rate τ , so the probability that a given infective contacts a given neighbour before the infective recovers is $\tau/(\gamma + \tau)$. Any contacted susceptible immediately becomes infective and may transmit the infection to any of its neighbouring susceptibles; i.e. there is no latent period. All the infectious periods and Poisson processes governing transmission of infection are mutually independent. The epidemic ends as soon as there is no infective present in the network.

2.2 The effective degree model

Ball and Neal [4] introduced an ‘effective-degree’ construction of the above epidemic, in which the network is constructed as the epidemic progresses. The process starts with some individuals infective and the remaining individuals susceptible, but with none of the stubs paired up. For $i = 1, 2, \dots, N$, the effective degree of individual i is initially D_i . Infected individuals transmit infection by pairing their stubs with stubs attached to susceptible individuals in the following fashion. An infected individual makes infectious contacts down its unpaired stubs independently at rate τ and is removed at rate γ . When an infective, individual i say, transmits infection down a stub that stub is paired with a stub (attached to individual j , say) chosen independently and uniformly at random from all the unpaired stubs, to form an edge. The effective degrees of individuals i and j are

both reduced by 1. (If $i = j$ then the effective degree of individual i is reduced by 2 but this will not significantly affect the dynamics for large populations since the probability of it happening is $O(N^{-1})$). If individual j is susceptible then it becomes infective and can transmit infection down any of its unattached stubs. As before, the epidemic ends as soon as there is no infective present. The network is then typically only partially constructed but that does not matter if interest is focussed on properties of the epidemic. In the original formulation of Ball and Neal [4], when an infective recovers its unpaired stubs, if any, were paired with stubs chosen uniformly at random without replacement from the set of unpaired stubs but that is unnecessary; the stubs from such an infective can simply be left in the set of unpaired stubs.

Ball and Neal [4] used the above approach to derive a system of ordinary differential equations that describes the deterministic limit of the epidemic model as the population size $N \rightarrow \infty$. A much simpler (equivalent) system of only 4 ordinary differential equations was obtained by Volz [31] and subsequently shown by Miller [23] to be essentially one-dimensional (the 4 ODEs were also shown by House and Keeling [20] to be a special case of the much higher dimensional pair approximation model of Eames and Keeling [14], in which the degree structure is explicit). Fully rigorous proofs of convergence in probability of the scaled stochastic model to the deterministic limit are given by Decreusefond *et al.* [11], Bohman and Piccollelli [7], Barbour and Reinart [5] and Janson *et al.* [21].

2.3 Approximating multitype branching process

Suppose that the size N of the network is large and the initial number of infectives is small. Then during the early stages of an epidemic it is very likely that the effective degree of a newly-infected individual, i say, is given by $D_i - 1$. It follows that the early stages of such an epidemic can be approximated by a branching process in which each newly-infected individual has their “full” effective degree (i.e. their actual degree minus one for the stub that is paired with their infector). This approximation can be made fully rigorous by considering a sequence of epidemics, indexed by N , and using a coupling argument; see e.g. Ball and Neal [4]. Moreover, the branching process is an upper bound for the epidemic process, so, as in the proof of Ball and Donnelly [3], Theorem 3, it is straightforward to use the dominated convergence theorem to prove convergence of moments of the number of infectives in the epidemic process to corresponding moments of the branching process as $N \rightarrow \infty$.

The limiting branching process may be described by a continuous-time multitype Markov branching process, with the type of an infective corresponding to its effective degree. Let \tilde{D} denote the (size-biased) degree of a typical neighbour of a typical individual in the network and let $\tilde{d}_k = \mathbb{P}(\tilde{D} = k)$ ($k = 1, 2, \dots$). Then $\tilde{d}_k = \mu_D^{-1} k d_k$, where $\mu_D = \mathbb{E}[D]$ (which we assume to be finite), since when a stub is paired it is k times as likely to be paired with a stub from a given individual having degree k than it is with a stub from a given individual having degree 1. Under the branching process approximation, the effective degree of a newly infected individual is distributed according to $\tilde{D} - 1$, since one of that individual’s stubs is ‘used up’ when it is infected. Note for future reference that $\mu_{\tilde{D}} = \mu_D^{-1} \mu_{D^2}$ and, more generally, $\mu_{f(\tilde{D})} = \mu_D^{-1} \mu_{Df(D)}$ for any real-valued function f . (For a random variable, X say, we use μ_X to denote its expectation $\mathbb{E}[X]$. Thus, for example, $\mu_{Df(D)} = \mathbb{E}[Df(D)]$.)

We now assume that there is a maximum degree k_{\max} , i.e. that $d_k = 0$ for all $k > k_{\max}$. We show later that almost all of our results extend to the case of no maximum degree size, subject to suitable moment conditions on D . Thus the type space for the branching

process is $\mathcal{K} = \{0, 1, \dots, k_{\max}\}$. Note that only initial infectives can have type k_{\max} . For $k \in \mathcal{K}$, an individual of type k dies if either its infectious period comes to an end or it transmits infection down one of its unattached stubs, whichever happens first. If the former happens first then the individual has no offspring, otherwise it has two offspring, namely an individual of type $k - 1$ and an individual whose type is distributed according to $\tilde{D} - 1$. Note that an individual of type 0 necessarily has no offspring when it dies. Thus, for $k \in \mathcal{K}$, the life time of an individual of type k has an exponential distribution with rate

$$\omega_k = \gamma + \tau k, \quad (2.1)$$

and when it dies its offspring is distributed as follows (recall that $\tilde{d}_{k_{\max}+1} = 0$):

$$\begin{aligned} \mathbb{P}(\text{Offspring} = \emptyset) &= \frac{\gamma}{\gamma + \tau k}, \\ \mathbb{P}(\text{Offspring} = \{k - 1, l\}) &= \frac{\tau k \tilde{d}_{l+1}}{\gamma + \tau k} \quad (l = 0, 1, \dots, k_{\max} - 1). \end{aligned} \quad (2.2)$$

The joint probability-generating function (PGF) for offspring of a type- k individual is therefore

$$P_k(\mathbf{s}) = \frac{1}{\omega_k} \left(\gamma + \tau k s_{k-1} \sum_{l=0}^{k_{\max}-1} \tilde{d}_{l+1} s_l \right),$$

where $\mathbf{s} = (s_k)$. In general we will write $\mathbf{v} = (v_k) = (v_1, \dots, v_{k_{\max}})^\top$ for a column vector in $\mathbb{R}^{k_{\max}+1}$, where $^\top$ denotes transpose. For $k \in \mathcal{K}$, let $\partial P_k(\mathbf{s})$ be the column vector whose i -th element is $\frac{\partial P_k(\mathbf{s})}{\partial s_i}$ and let $\partial^2 P_k(\mathbf{s})$ be the matrix whose (i, j) -th element is $\frac{\partial^2 P_k(\mathbf{s})}{\partial s_i \partial s_j}$. We note for future reference that, for $i, j, k \in \mathcal{K}$,

$$\begin{aligned} [\partial P_k(\mathbf{1})]_i &= \frac{\tau k}{\gamma + \tau k} \left(\tilde{d}_{i+1} + \delta_{k-1, i} \right) \\ [\partial^2 P_k(\mathbf{1})]_{i, j} &= \frac{\tau k}{\gamma + \tau k} \left(\tilde{d}_{i+1} + \delta_{k-1, j} + \tilde{d}_{j+1} + \delta_{k-1, i} \right), \end{aligned} \quad (2.3)$$

where $\mathbf{1}$ is the length- $(k_{\max} + 1)$ column vector of ones.

For $t \geq 0$, let $\mathbf{Z}(t) = (Z_i(t))$, where $Z_i(t)$ denotes the number of individuals of type i alive at time t , and let $Z(t) = Z_0(t) + Z_1(t) + \dots + Z_{k_{\max}}(t) = \mathbf{1}^\top \mathbf{Z}(t)$ denote the total number individuals alive at time t . For $k \in \mathcal{K}$, we use the notation $\{\mathbf{Z}^{(k)}(t) : t \geq 0\}$, where $\mathbf{Z}^{(k)}(t) = (Z_i^{(k)}(t))$, to denote a process starting with a single individual, whose type is k , at time 0 (i.e. where $Z_i^{(k)}(0) = \delta_{i, k}$, $i \in \mathcal{K}$). Further, $Z^{(k)}(t) = \mathbf{1}^\top \mathbf{Z}^{(k)}(t)$ denotes the total number of individuals at time t in such a process.

3 Behaviour of means

For $t \geq 0$ and $i, j, k \in \mathcal{K}$, let

$$\begin{aligned} \mathbf{M}(t) &= [m_{i, j}(t)], \text{ where } m_{i, j}(t) = \mathbb{E} \left[Z_j^{(i)}(t) \right], \\ \mathbf{m}^{(k)}(t) &= \mathbb{E} \left[\mathbf{Z}^{(k)}(t) \right] = \mathbf{u}_k^\top \mathbf{M}(t), \\ m^{(k)}(t) &= \mathbb{E} \left[Z^{(k)}(t) \right] = \mathbf{1}^\top \mathbf{m}^{(k)}(t), \end{aligned}$$

where \mathbf{u}_k is a length- $(k_{\max} + 1)$ column vector with k -th element equal to 1 and other elements equal to 0. A standard argument using the Kolmogorov forward equation (see e.g. Dormer *et al.* [12], Section 7 and recall that $\tilde{d}_{k_{\max}+1} = 0$) then yields that

$$\frac{d}{dt}\mathbf{M}(t) = \mathbf{M}(t)\mathbf{\Omega} , \quad \mathbf{M}(0) = \mathbf{I} , \quad (3.1)$$

where \mathbf{I} denotes the $(k_{\max} + 1) \times (k_{\max} + 1)$ identity matrix and $\mathbf{\Omega} = [\Omega_{l,k}]$ is the $(k_{\max} + 1) \times (k_{\max} + 1)$ matrix with elements given by

$$\Omega_{l,k} = \tau l \left(\tilde{d}_{k+1} + \delta_{l,k+1} \right) - (\gamma + \tau l) \delta_{l,k} \quad (l, k \in \mathcal{K}) .$$

The solution to (3.1) is then straightforwardly given by

$$\mathbf{M}(t) = e^{\mathbf{\Omega}t} = \sum_{l=0}^{\infty} \frac{t^l \mathbf{\Omega}^l}{l!} . \quad (3.2)$$

We show in Appendix A that the eigenvalues of $\mathbf{\Omega}$ are

$$\lambda_i = \begin{cases} -\gamma - i\tau & \text{for } i \in \{0, 2, 3, \dots, k_{\max}\} , \\ \tau((\sum_{l=0}^{k_{\max}} l \tilde{d}_{l+1}) - 1) - \gamma & \text{for } i = 1 . \end{cases} \quad (3.3)$$

We denote the dominant eigenvalue, λ_1 , by r , so

$$r = \tau((\sum_{l=0}^{k_{\max}} l \tilde{d}_{l+1}) - 1) - \gamma = \tau \mu_{\tilde{D}-2} - \gamma . \quad (3.4)$$

If $r \leq 0$, the branching process $\{\mathbf{Z}(t) : t \geq 0\}$ goes extinct almost surely. If $r > 0$, then r gives the asymptotic exponential growth rate of $\{Z(t) : t \geq 0\}$ (and also of $\{Z_i(t) : t \geq 0\}$ for $i \in \mathcal{K} \setminus \{k_{\max}\}$) when $\{\mathbf{Z}(t) : t \geq 0\}$ does not go extinct.

For $i \in \mathcal{K}$, let $\mathbf{w}_i^\top = (w_{i,k})$ be a left eigenvector of $\mathbf{\Omega}$ corresponding to the eigenvalue λ_i , so

$$\mathbf{w}_i^\top \mathbf{\Omega} = \lambda_i \mathbf{w}_i^\top . \quad (3.5)$$

The Perron-Frobenius theory implies that \mathbf{w}_1 can be chosen so that all of its elements are positive and $\mathbf{w}_1^\top \mathbf{1} = 1$. The left-eigenvector \mathbf{w}_1 then yields a probability distribution which gives the asymptotic relative frequencies of the different types, as $t \rightarrow \infty$, when $\{\mathbf{Z}(t) : t \geq 0\}$ does not go extinct.

Expanding (3.5) in components yields

$$\sum_{l=0}^{k_{\max}} w_{1,l} \left(\tau l \left(\tilde{d}_{k+1} + \delta_{l,k+1} \right) - (\gamma + \tau l) \delta_{l,k} \right) = r w_{1,k} \quad (k \in \mathcal{K}) . \quad (3.6)$$

Let $w(s) = \sum_{l=0}^{k_{\max}} s^l w_{1,l}$ ($s \geq 0$) denote the (probability-)generating function of \mathbf{w}_1 . Multiplying (3.6) by s^k and summing over k yields

$$\tau f_{\tilde{D}-1}(s) \mu_W + \tau(1-s)w'(s) = (r + \gamma)w(s) , \quad (3.7)$$

where $f_{\tilde{D}-1}(s) = \sum_{k=1}^{k_{\max}} \tilde{d}_k s^{k-1}$ is the PGF of $\tilde{D} - 1$ and $\mu_W = \sum_{k=0}^{k_{\max}} k w_{1,k}$ is the mean of the distribution \mathbf{w}_1 . Setting $s = 1$ in (3.7) and using (3.4) yields

$$\mu_W = \frac{r + \gamma}{\tau} = \mu_{\tilde{D}-2} . \quad (3.8)$$

For $i, k \in \mathbb{Z}^+$, let $k_{[i]} = k(k-1) \dots (k-i+1)$ denote a falling factorial, with the convention $k_{[0]} = 0$. For $i = 0, 1, \dots$, let $\mu_W^{[i]} = \sum_{k=0}^{k_{\max}} k_{[i]} w_{1,k}$ be the i th factorial-moment of the distribution \mathbf{w}_1 , so $\mu_W^{[0]} = 1$ and $\mu_W^{[1]} = \mu_W$. Note that $\mu_W^{[i]} = w^{(i)}(1)$ ($i = 0, 1, \dots$), where $w^{(i)}(s)$ denotes the i -th derivative of $w(s)$. Repeated differentiation of (3.7) yields

$$\mu_W^{[i]} = \frac{\mu_{\tilde{D}-2}}{\mu_{\tilde{D}-2+i}} \mu_{\tilde{D}-1}^{[i]}, \quad (3.9)$$

where $\mu_{\tilde{D}-1}^{[i]}$ is the i -th factorial-moment of $\tilde{D} - 1$. Note that $\mu_{\tilde{D}-1}^{[i]} = 0$ for $i \geq k_{\max}$. It then follows, using the inversion formula which expresses the probability mass function of a non-negative integer-valued random variable in terms of its factorial-moments (see e.g. Daley and Vere-Jones [9], page 117), that

$$w_{1,k} = \begin{cases} \sum_{i=k}^{k_{\max}-1} (-1)^{i-k} \binom{i}{k} \frac{\mu_{\tilde{D}-2} \mu_{\tilde{D}-1}^{[i]}}{i! \mu_{\tilde{D}-2+i}} & \text{if } k = 0, 1, \dots, k_{\max} - 1, \\ 0 & \text{if } k = k_{\max}. \end{cases} \quad (3.10)$$

Observe that $w_{1,k_{\max}} = 0$ since only initial infectives can have type k_{\max} . Observe also that \mathbf{w}_1 is determined just by the degree distribution of the network and is invariant to the epidemic parameters γ and τ .

For $t \geq 0$, let $\mathbf{m}(t) = (m^{(i)}(t)) = \mathbf{M}(t)\mathbf{1}$. Thus the k -th element of $\mathbf{m}(t)$ contains the mean total population size at time t given that the process starts with a single individual whose type is k . We derive a simple expression for $\mathbf{m}(t)$. The following proposition is useful.

Proposition 1. *For a matrix \mathbf{M} and vectors \mathbf{x}, \mathbf{y} such that $\mathbf{M}\mathbf{x} = a\mathbf{x} + b\mathbf{y}$ and $\mathbf{M}\mathbf{y} = c\mathbf{y}$, where a, b and c are scalars satisfying $a \neq c$,*

$$e^{\mathbf{M}t}\mathbf{x} = e^{at}\mathbf{x} + \frac{b}{a-c} (e^{at} - e^{ct})\mathbf{y} \quad \text{and} \quad e^{\mathbf{M}t}\mathbf{y} = e^{ct}\mathbf{y}. \quad (3.11)$$

Proof. The second identity follows straightforwardly from the definition of the matrix exponential and the fact that \mathbf{y} is a right eigenvector with eigenvalue c . For the first identity,

$$\begin{aligned} e^{\mathbf{M}t}\mathbf{x} &= \sum_{i=0}^{\infty} \frac{1}{i!} (\mathbf{M}t)^i \mathbf{x} \\ &= \sum_{i=0}^{\infty} \frac{t^i}{i!} \left(a^i \mathbf{x} + \sum_{j=0}^{i-1} b c^{i-1-j} a^j \mathbf{y} \right) \\ &= e^{at}\mathbf{x} + \sum_{i=0}^{\infty} \frac{t^i}{i!} b c^{i-1} \frac{\left(\frac{a}{c}\right)^i - 1}{\left(\frac{a}{c}\right) - 1} \mathbf{y} \\ &= e^{at}\mathbf{x} + \frac{b}{a-c} (e^{at} - e^{ct})\mathbf{y}. \end{aligned} \quad (3.12)$$

□

Let $\mathbf{n} = (0, 1, \dots, k_{\max})^\top$. Observe that

$$\Omega \mathbf{1} = \tau \mathbf{n} - \gamma \mathbf{1} \quad \text{and} \quad \Omega \mathbf{n} = r \mathbf{n}, \quad (3.13)$$

so using Proposition 1, and recalling from (3.4) that $r + \gamma = \mu_{\tilde{D}-2}\tau$, we have

$$e^{\Omega x} \mathbf{1} = \mu_{\tilde{D}-2}^{-1} (e^{rx} - e^{-\gamma x}) \mathbf{n} + e^{-\gamma x} \mathbf{1} \quad \text{and} \quad e^{\Omega x} \mathbf{n} = e^{rx} \mathbf{n} . \quad (3.14)$$

Thus,

$$\mathbf{m}(t) = \mu_{\tilde{D}-2}^{-1} (e^{rt} - e^{-\gamma t}) \mathbf{n} + e^{-\gamma t} \mathbf{1} . \quad (3.15)$$

and

$$\lim_{t \rightarrow \infty} e^{-rt} \mathbf{m}(t) = \mu_{\tilde{D}-2}^{-1} \mathbf{n} . \quad (3.16)$$

While it is well known that asymptotically the mean prevalence grows exponentially with rate constant r , i.e. that prevalence $\propto e^{rt}$, these results allow us to see from (3.15) that the rate of convergence to this asymptotic behaviour is $r + \gamma$, and from (3.16) that the constant of proportionality is the degree of the initially infected individual divided by $\mu_{\tilde{D}-2}$.

4 Variance

The variance in infectious prevalence was considered in Graham and House [16], but using the diffusion limit and an argument about the neighbourhood around an infective. A branching process limit lets us be more explicit. For $k \in \mathcal{K}$, let \mathbf{u}_k denote the length- $(k_{\max} + 1)$ column vector whose k -th element is 1 and all other elements are 0, so $\mathbf{u}_k = (\delta_{i,k})$. A matrix integrating factor argument gives

$$\begin{aligned} \mathbf{V}^{(k)}(t) &= \mathbb{E} \left[\mathbf{Z}^{(k)}(t) \mathbf{Z}^{(k)}(t)^\top \right] - \mathbb{E} \left[\mathbf{Z}^{(k)}(t) \right] \mathbb{E} \left[\mathbf{Z}^{(k)}(t)^\top \right] \\ &= \int_0^t e^{\Omega^\top(t-u)} \mathbf{B}_k(u) e^{\Omega(t-u)} du , \end{aligned} \quad (4.1)$$

where

$$\begin{aligned} \mathbf{B}_k(t) &= \sum_{l=0}^{k_{\max}} (e^{\Omega t})_{k,l} \mathbf{C}_l , \\ \mathbf{C}_k &= \omega_k \left(\partial^2 P_k(\mathbf{1}) + \text{diag}(\mathbf{f}_k) - \mathbf{u}_k \mathbf{f}_k^\top - \mathbf{f}_k \mathbf{u}_k^\top + \mathbf{u}_k \mathbf{u}_k^\top \right) , \\ \mathbf{f}_k &= \partial P_k(\mathbf{1}) . \end{aligned}$$

See Dorman *et al.* [12], Section 9, and also Athreya and Ney [1], page 203, for details. Note that there appears to be a small error in the latter – in the expression for $b_{jk}^{(i)}$ on page 203 of [1], $\delta_{jk} - b_{ij}\delta_{ik} - b_{ik}\delta_{ij} + \delta_{ij}\delta_{ik}$ should be replaced by $b_{ij}\delta_{jk} - b_{ik}\delta_{ij} - b_{ij}\delta_{ik}$.

For $t \geq 0$ and $k \in \mathcal{K}$, let $v^{(k)}(t)$ denote the variance of the total population size at time t given that the process starts with a single individual, whose type is k . Then $v^{(k)}(t) = \mathbf{1}^\top \mathbf{V}^{(k)}(t) \mathbf{1}$ and it follows using (4.1) that

$$\begin{aligned} v^{(k)}(t) &= \int_0^t (e^{\Omega(t-u)} \mathbf{1})^\top \mathbf{B}_k(u) (e^{\Omega(t-u)} \mathbf{1}) du \\ &= \int_0^t \left(\mu_{\tilde{D}-2}^{-1} (e^{r(t-u)} - e^{-\gamma(t-u)}) \right)^2 \mathbf{n}^\top \mathbf{B}_k(u) \mathbf{n} du \\ &\quad + 2 \int_0^t \mu_{\tilde{D}-2}^{-1} (e^{r(t-u)} - e^{-\gamma(t-u)}) e^{-\gamma(t-u)} \mathbf{1}^\top \mathbf{B}_k(u) \mathbf{n} du \\ &\quad + \int_0^t e^{-2\gamma(t-u)} \mathbf{1}^\top \mathbf{B}_k(u) \mathbf{1} du , \end{aligned} \quad (4.2)$$

where we have used the first equation in (3.14) in deriving the last line. This quantity will turn out to have an exact but rather complex closed-form solution, so we begin by building up some intermediate results. First note, using (2.3), that for $k \in \mathcal{K}$,

$$\begin{aligned} \mathbf{1}^\top \mathbf{C}_k \mathbf{1} &= \tau k + \gamma, \\ \mathbf{1}^\top \mathbf{C}_k \mathbf{n} &= \mathbf{n}^\top \mathbf{C}_k \mathbf{1} = (r + 2\gamma)k, \\ \mathbf{n}^\top \mathbf{C}_k \mathbf{n} &= \tau \mu_{(\tilde{D}-2)^2} k + \gamma k^2. \end{aligned} \quad (4.3)$$

Now let \mathbf{c}_{11} be a column vector whose k -th element is $\mathbf{1}^\top \mathbf{C}_k \mathbf{1}$, and define \mathbf{c}_{1n} and \mathbf{c}_{nn} similarly, using $\mathbf{1}^\top \mathbf{C}_k \mathbf{n}$ and $\mathbf{n}^\top \mathbf{C}_k \mathbf{n}$, respectively. Noting that $r + 2\gamma = \gamma + \tau \mu_{\tilde{D}-2}$, in this more compact notation (4.3) becomes

$$\begin{aligned} \mathbf{c}_{11} &= \tau \mathbf{n} + \gamma \mathbf{1}, \\ \mathbf{c}_{1n} &= (\gamma + \tau \mu_{\tilde{D}-2}) \mathbf{n}, \\ \mathbf{c}_{nn} &= \tau \mu_{(\tilde{D}-2)^2} \mathbf{n} + \gamma \mathbf{n}_2, \end{aligned} \quad (4.4)$$

where $\mathbf{n}_2 = (0^2, 1^2, \dots, k_{\max}^2)^\top$. For $t \geq 0$, let $\mathbf{v}(t) = (v^{(i)}(t))$. Then (4.2) yields

$$\begin{aligned} \mathbf{v}(t) &= \int_0^t \left(\mu_{\tilde{D}-2}^{-1} \left(e^{r(t-u)} - e^{-\gamma(t-u)} \right) \right)^2 e^{\Omega u} \mathbf{c}_{nn} \, du \\ &\quad + 2 \int_0^t \mu_{\tilde{D}-2}^{-1} \left(e^{r(t-u)} - e^{-\gamma(t-u)} \right) e^{-\gamma(t-u)} e^{\Omega u} \mathbf{c}_{1n} \, du \\ &\quad + \int_0^t e^{-2\gamma(t-u)} e^{\Omega u} \mathbf{c}_{11} \, du. \end{aligned} \quad (4.5)$$

Now

$$\Omega \mathbf{n}_2 = \tau \mu_{(\tilde{D}-1)^2+1} \mathbf{n} - (2\tau + \gamma) \mathbf{n}_2, \quad (4.6)$$

so, using Proposition 1,

$$e^{\Omega u} \mathbf{n}_2 = \mu_{\tilde{D}}^{-1} \mu_{(\tilde{D}-1)^2+1} \left(e^{ru} - e^{-(2\tau+\gamma)u} \right) \mathbf{n} + e^{-(2\tau+\gamma)u} \mathbf{n}_2. \quad (4.7)$$

Hence, using also (3.14),

$$\begin{aligned} e^{\Omega u} \mathbf{c}_{11} &= \mu_{\tilde{D}-2}^{-1} (\gamma + \tau \mu_{\tilde{D}-2}) e^{ru} \mathbf{n} + \gamma e^{-\gamma u} \left(\mathbf{1} - \mu_{\tilde{D}-2}^{-1} \mathbf{n} \right), \\ e^{\Omega u} \mathbf{c}_{1n} &= (\gamma + \tau \mu_{\tilde{D}-2}) e^{ru} \mathbf{n}, \\ e^{\Omega u} \mathbf{c}_{nn} &= \left(\tau \mu_{(\tilde{D}-2)^2} + \gamma \mu_{\tilde{D}}^{-1} \mu_{(\tilde{D}-1)^2+1} \right) e^{ru} \mathbf{n} + \gamma e^{-(2\tau+\gamma)u} \left(\mathbf{n}_2 - \mu_{\tilde{D}}^{-1} \mu_{(\tilde{D}-1)^2+1} \mathbf{n} \right). \end{aligned} \quad (4.8)$$

Let

$$\begin{aligned}
I_1(t) &= \int_0^t e^{-2\gamma(t-u)} e^{ru} du = \frac{e^{rt} - e^{-2\gamma t}}{r + 2\gamma} , \\
I_2(t) &= \int_0^t e^{-2\gamma(t-u)} e^{-\gamma u} du = \frac{e^{-\gamma t} (1 - e^{-\gamma t})}{\gamma} , \\
I_3(t) &= \int_0^t \left(e^{r(t-u)} - e^{-\gamma(t-u)} \right) e^{-\gamma(t-u)} e^{ru} du = \frac{e^{rt} (1 - e^{-\gamma t})}{\gamma} - I_1(t) , \\
I_4(t) &= \int_0^t \left(e^{r(t-u)} - e^{-\gamma(t-u)} \right)^2 e^{ru} du = \frac{e^{rt} (e^{rt} - 1)}{r} - 2I_3(t) - I_1(t) , \\
I_5(t) &= \int_0^t \left(e^{r(t-u)} - e^{-\gamma(t-u)} \right)^2 e^{-(2\tau+\gamma)u} du = \frac{e^{2rt} - e^{-(\gamma+2\tau)t}}{2r + 2\tau + \gamma} \\
&\quad - 2 \frac{e^{-\gamma t} (e^{rt} - e^{-2\tau t})}{r + 2\tau} + \frac{e^{-\gamma t} (e^{-2\tau t} - e^{-\gamma t})}{\gamma - 2\tau} ,
\end{aligned} \tag{4.9}$$

where, if $\gamma = 2\tau$, the final term in $I_5(t)$ is replaced by $te^{-2\gamma t}$. Substituting (4.8) into (4.5) and using (4.9) yields

$$\mathbf{v}(t) = \alpha_0(t) \mathbf{1} + \alpha_1(t) \mathbf{n} + \alpha_2(t) \mathbf{n}_2 , \tag{4.10}$$

where

$$\begin{aligned}
\alpha_0(t) &= \gamma I_2(t) , \\
\alpha_1(t) &= \gamma \mu_{\tilde{D}-2}^{-1} \left[I_1(t) - I_2(t) + 2I_3(t) + \mu_{\tilde{D}-2}^{-1} \mu_{\tilde{D}}^{-1} \mu_{(\tilde{D}-1)^2+1} (I_4(t) - I_5(t)) \right] \\
&\quad + \tau \left[I_1(t) + 2I_3(t) + \mu_{\tilde{D}-2}^{-2} \mu_{(\tilde{D}-2)^2} I_4(t) \right] , \\
\alpha_2(t) &= \gamma \mu_{\tilde{D}-2}^{-2} I_5(t) .
\end{aligned} \tag{4.11}$$

The equation (4.10) leads to a rather complex expression for $\mathbf{v}(t)$ in terms of elementary functions. However, its asymptotic form as $t \rightarrow \infty$ is much simpler. Note that $\lim_{t \rightarrow \infty} e^{-2rt} I_k(t) = 0$, for $k = 1, 2, 3$, $\lim_{t \rightarrow \infty} e^{-2rt} I_4(t) = \frac{1}{r}$ and $\lim_{t \rightarrow \infty} e^{-2rt} I_5(t) = \frac{1}{2r+2\tau+\gamma} = \frac{1}{2\mu_{\tilde{D}-1}\tau - \gamma}$. Substituting these limits into (4.10) and (4.11) yields

$$\lim_{t \rightarrow \infty} e^{-2rt} \mathbf{v}(t) = \frac{1}{\mu_{\tilde{D}-2}^2 (2\mu_{\tilde{D}-1}\tau - \gamma)} \left[\frac{2\tau \mu_{\tilde{D}-1} (\mu_{(\tilde{D}-2)^2}\tau + \gamma)}{r} \mathbf{n} + \gamma \mathbf{n}_2 \right] . \tag{4.12}$$

Note that both the asymptotic and exact expressions for $\mathbf{v}(t)$ depend on the degree distribution D only through its first three moments.

It follows from (3.16) and (4.12) that, for $k \in \mathcal{K}$,

$$\lim_{t \rightarrow \infty} \frac{\text{var} (Z^{(k)}(t))}{\mathbb{E} [Z^{(k)}(t)]^2} = \frac{1}{2\mu_{\tilde{D}-1}\tau - \gamma} \left[\gamma + \frac{2\tau \mu_{\tilde{D}-1} (\mu_{(\tilde{D}-2)^2}\tau + \gamma)}{kr} \right] . \tag{4.13}$$

We note two features of these results. First, the equations (4.10) and (4.11) involve many rates that are linear combinations of r , τ and γ , with the dominant being $2r$ and the subdominant being r . This leads to complex real-time behaviour as the system approaches

its asymptotic limit. In the diffusion limit, only the dominant and subdominant rates are present, leading to the same overall rate of convergence r , but other rates are not present [16]. Secondly, the dependence of the variance on initial conditions is not simple proportionality, meaning that (4.12) contains terms proportional to both \mathbf{n} and \mathbf{n}_2 and the right-hand side of (4.13) depends on k .

5 Covariance function

For $t, s \geq 0$ and $k \in \mathcal{K}$, let $\sigma^{(k)}(t, s) = \text{cov}(Z^{(k)}(t), Z^{(k)}(s))$ denote the covariance of the total population sizes at times t and s , given that the process starts with a single individual, whose type is k . Assume without loss of generality that $t \leq s$. Then

$$\begin{aligned} \sigma^{(k)}(t, s) &= \mathbb{E} \left[\text{cov} \left(Z^{(k)}(t), Z^{(k)}(s) | \mathbf{Z}^{(k)}(t) \right) \right] \\ &\quad + \text{cov} \left(\mathbb{E} \left[Z^{(k)}(t) | \mathbf{Z}^{(k)}(t) \right], \mathbb{E} \left[Z^{(k)}(s) | \mathbf{Z}^{(k)}(t) \right] \right). \end{aligned} \quad (5.1)$$

The first term on the right hand side of (5.1) is zero, since $Z^{(k)}(t)$ is non-random given $\mathbf{Z}^{(k)}(t)$. Now, $\mathbb{E} [Z^{(k)}(t) | \mathbf{Z}^{(k)}(t)] = \mathbf{1}^\top \mathbf{Z}^{(k)}(t)$ and $\mathbb{E} [Z^{(k)}(s) | \mathbf{Z}^{(k)}(t)] = \mathbf{Z}^{(k)}(t)^\top \mathbf{M}(s-t) \mathbf{1}$, so

$$\begin{aligned} \sigma^{(k)}(t, s) &= \text{cov} \left(\mathbf{1}^\top \mathbf{Z}^{(k)}(t), \mathbf{Z}^{(k)}(t)^\top \mathbf{M}(s-t) \mathbf{1} \right) \\ &= \mathbf{1}^\top \mathbf{V}^{(k)}(t) \mathbf{M}(s-t) \mathbf{1} \\ &= \mu_{\bar{D}-2}^{-1} \left(e^{r(s-t)} - e^{-\gamma(s-t)} \right) \mathbf{1}^\top \mathbf{V}^{(k)}(t) \mathbf{n} + e^{-\gamma(s-t)} \mathbf{1}^\top \mathbf{V}^{(k)}(t) \mathbf{1}, \end{aligned} \quad (5.2)$$

using (3.2) and the first equation in (3.14). Now $\mathbf{1}^\top \mathbf{V}^{(k)}(t) \mathbf{1} = v^k(t)$, and letting $\mathbf{u}(t)$ be the column vector whose k -th element is $\mathbf{1}^\top \mathbf{V}^{(k)}(t) \mathbf{n}$ and arguing as in the derivation of (4.5) yields

$$\begin{aligned} \mathbf{u}(t) &= \int_0^t \mu_{\bar{D}-2}^{-1} \left(e^{r(t-u)} - e^{-\gamma(t-u)} \right) e^{r(t-u)} e^{\mathbf{\Omega}u} \mathbf{c}_{nn} \, du \\ &\quad + \int_0^t e^{-\gamma(t-u)} e^{r(t-u)} e^{\mathbf{\Omega}u} \mathbf{c}_{1n} \, du. \end{aligned} \quad (5.3)$$

Let

$$\begin{aligned} I_6(t) &= \int_0^t e^{-\gamma(t-u)} e^{r(t-u)} e^{ru} \, du = \frac{e^{rt} (1 - e^{-\gamma t})}{\gamma}, \\ I_7(t) &= \int_0^t \left(e^{r(t-u)} - e^{-\gamma(t-u)} \right) e^{r(t-u)} e^{ru} \, du = \frac{e^{rt} (e^{rt} - 1)}{r} - I_6(t), \\ I_8(t) &= \int_0^t \left(e^{r(t-u)} - e^{-\gamma(t-u)} \right) e^{r(t-u)} e^{-(2\tau+\gamma)u} \, du = \frac{e^{2rt} - e^{-(\gamma+2\tau)t}}{2r+2\tau+\gamma} - \frac{e^{-\gamma t} (e^{rt} - e^{-2\tau t})}{r+2\tau}. \end{aligned} \quad (5.4)$$

Then, using (4.8) yields that

$$\mathbf{u}(t) = \beta_1(t) \mathbf{n} + \beta_2(t) \mathbf{n}_2, \quad (5.5)$$

where

$$\begin{aligned} \beta_1(t) &= \gamma \mu_{\bar{D}-2}^{-1} \left[\mu_{\bar{D}}^{-1} \mu_{(\bar{D}-1)^2+1} (I_7(t) - I_8(t)) + \mu_{\bar{D}-2} I_6(t) \right] \\ &\quad + \tau \mu_{\bar{D}-2}^{-1} \left[\mu_{(\bar{D}-2)^2} I_7(t) + \mu_{\bar{D}-2}^2 I_6(t) \right], \\ \beta_2(t) &= \gamma \mu_{\bar{D}-2}^{-1} I_8(t). \end{aligned} \quad (5.6)$$

For $t, s \geq 0$, let $\sigma(t, s) = (\sigma^{(0)}(t), \sigma^{(1)}(t), \dots, \sigma^{(k_{\max})}(t))^{\top}$. It follows from (5.2) and (5.5) that, for $0 \leq t \leq s$,

$$\sigma(t, s) = \mu_{\tilde{D}-2}^{-1} \left(e^{r(s-t)} - e^{-\gamma(s-t)} \right) \mathbf{u}(t) + e^{-\gamma(s-t)} \mathbf{v}(t) . \quad (5.7)$$

Now, $\lim_{t \rightarrow \infty} e^{-2rt} I_6(t) = 0$, $\lim_{t \rightarrow \infty} e^{-2rt} I_7(t) = \frac{1}{r}$ and $\lim_{t \rightarrow \infty} e^{-2rt} I_8(t) = \frac{1}{2\mu_{\tilde{D}-1}\tau - \gamma}$. Substituting these limits into (5.7) yields that, for any $s \geq 0$,

$$\lim_{t \rightarrow \infty} e^{-2rt} \sigma(t, t+s) = e^{rs} \lim_{t \rightarrow \infty} e^{-2rt} \mathbf{v}(t) . \quad (5.8)$$

It follows that, for $k \in \mathcal{K}$ and $s > 0$,

$$\lim_{t \rightarrow \infty} \text{corr} \left(Z^{(k)}(t), Z^{(k)}(t+s) \right) = 1 ,$$

where corr denotes correlation. This is not surprising since it is well known that

$$e^{-rt} Z^{(k)}(t) \xrightarrow{\text{a.s.}} W^{(k)} \quad \text{as } t \rightarrow \infty ,$$

where $\xrightarrow{\text{a.s.}}$ denotes almost sure convergence (i.e. convergence with probability 1) and $W^{(k)}$ is a non-negative random which satisfies $W^{(k)} = 0$ if and only if the branching process goes extinct; see e.g. Athreya and Ney [1], Theorem V.7.2.

6 Unbounded degree distributions

The above results have all assumed that there is a maximum degree k_{\max} . Suppose that is not the case, so the branching process has countably many types. For $t \geq 0$, let $\mathbf{Z}(t) = (Z_0(t), Z_1(t), \dots)^{\top}$, where $Z_i(t)$ denotes the number of individuals of type i alive at time t , and let $Z(t) = \sum_{i=0}^{\infty} Z_i(t)$ denote the total number individuals alive at time t . (For ease of notation we drop explicit reference to the type of the initial individual.) For $k_{\max} = 1, 2, \dots$, let $\{\mathbf{Z}(t, k_{\max}) : t \geq 0\}$ denote the branching process derived from $\{\mathbf{Z}(t) : t \geq 0\}$ by ignoring all individuals having type strictly greater than k_{\max} and any offspring of such individuals. For $t \geq 0$, let $Z(t, k_{\max}) = \mathbf{1}^{\top} \mathbf{Z}(t, k_{\max})$ be the total number of individuals alive in $\{\mathbf{Z}(t, k_{\max}) : t \geq 0\}$ at time t . Now, for any $t \geq 0$, $Z(t, k_{\max})$ is monotonically increasing in k_{\max} and converges almost surely to $Z(t)$ as $k_{\max} \rightarrow \infty$. Thus, by the monotone convergence theorem, $\mathbb{E}[Z(t)] = \lim_{k_{\max} \rightarrow \infty} \mathbb{E}[Z(t, k_{\max})]$.

The process $\{\mathbf{Z}(t, k_{\max}) : t \geq 0\}$ behaves like the branching process described in Section 2.3 but with infection rate τ replaced by $\tau(k_{\max}) = \tau \mathbb{P}(\tilde{D} \leq k_{\max} + 1)$, and size-biased degree distribution \tilde{D} replaced by $\tilde{D}(k_{\max})$, where

$$\mathbb{P}(\tilde{D}(k_{\max}) = k) = \begin{cases} \frac{\tilde{d}_k}{\mathbb{P}(\tilde{D} \leq k_{\max} + 1)} & \text{if } k = 1, 2, \dots, k_{\max} + 1 , \\ 0 & \text{if } k = k_{\max} + 2, k_{\max} + 3, \dots . \end{cases}$$

This is because contacts with individuals having degree strictly greater than $k_{\max} + 1$ are ignored, as they yield individuals with effective degree (and hence type) strictly greater than k_{\max} . Now $\tau(k_{\max}) \rightarrow \tau$ and $\mathbb{E}[\tilde{D}(k_{\max})] \rightarrow \mathbb{E}[\tilde{D}]$ as $k_{\max} \rightarrow \infty$, so the expression (3.15) for the mean total population size at time t continues to hold in the unbounded degree case, provided that $\mathbb{E}[\tilde{D}] < \infty$, or equivalently that $\mathbb{E}[D^2] < \infty$. A similar argument shows that the expressions for the variance of $Z(t)$ and the covariance of $Z(t)$ and

$Z(s)$, derived in Sections 4 and 5, respectively, continue to hold provided $\mathbb{E}[\tilde{D}^2] < \infty$, or equivalently $\mathbb{E}[D^3] < \infty$. In this context, we note that the weakest conditions obtained on the moments of the degree distribution for convergence of the mean of the full stochastic epidemic onto its deterministic limit are given by Janson *et al.* [21], who require uniform boundedness of the second moment of D .

7 Probability of extinction

For $t \geq 0$ and $k \in \mathcal{K}$, let $\pi_k(t) = \mathbb{P}(Z^{(k)}(t) = 0)$ be the probability that the branching process is extinct at time t given that it starts with one individual of type k . Then in general

$$\frac{d}{dt}\pi_k(t) = -\omega_k\pi_k(t) + \omega_k P_k(\boldsymbol{\pi}(t)) , \quad (7.1)$$

where $\boldsymbol{\pi}(t) = (\pi_i(t))$. For our specific model we have

$$\frac{d}{dt}\pi_k(t) = -(\gamma + \tau k)\pi_k(t) + \gamma + \tau k\pi_{k-1}(t) \sum_{l=0}^{k_{\max}-1} \tilde{d}_{l+1}\pi_l(t) . \quad (7.2)$$

These equations are not amenable to closed-form solution. Note, however, that studies of time to extinction for network epidemics – e.g. Holme [19] – have tended to be based on Monte Carlo methods, but (7.2) could provide a complementary approach that is numerically cheaper and more analytically tractable.

We will now consider three regimes in which asymptotic methods can be used to bound the real-time behaviour of the probabilities of extinction. In particular, we will see that early real-time behaviour is bounded by the death rates ω_k , while late-time behaviour is bounded by the asymptotic real-time growth rate r provided $r > -\gamma$.

7.1 Early behaviour

Now note that for $k \in \mathcal{K}$, $\pi_k(0) = 0$ and $\pi_k(t)$ is monotonically increasing with t . If we neglect the quadratic terms in (7.2) then, since these are only positive and increasing over time, we get a lower bound for the rate of convergence of the extinction probabilities; in particular we get

$$\pi_k(t) = \frac{\gamma}{\gamma + \tau k} \left(1 - e^{-(\gamma + \tau k)t}\right) + O(|\pi|^2) . \quad (7.3)$$

This tells us that the rate at which the early extinction probability starting with an individual of degree k converges faster than its death rate $\omega_k = \gamma + \tau k$ onto its final value. In the subcritical case, this final value will be 1 for all k , while for the supercritical case we expect different values for each k in the interval $(0, 1)$.

7.2 Late behaviour of the subcritical case

Suppose that $r < 0$, so the branching process is subcritical. For $t \geq 0$ and $k \in \mathbb{N}_0 = \{0, 1, \dots\}$, we will work with the probability of survival $q_k(t) = 1 - \pi_k(t) = \mathbb{P}(Z^{(k)}(t) > 0)$. Now $\mathbb{P}(Z^{(k)}(t) > 0) \leq \mathbb{E}[Z^{(k)}(t)]$, so using (3.15) a simple upper bound for $q_k(t)$, valid also in the unbounded degree setting using the results in Section 6, is

$$q_k(t) \leq k\mu_{\tilde{D}-2}^{-1} (e^{rt} - e^{-\gamma t}) + e^{-\gamma t} . \quad (7.4)$$

Note that $\mu_{\tilde{D}} < \infty$ is a necessary condition for $r < 0$. Under the stronger condition that $\mu_{\tilde{D}^2} < \infty$, Windridge [33] gives an exponential approximation, for large t , to a quantity closely related to $q_k(t)$. He assumes that what we call type-0 individuals are dead. For $k = 1, 2, \dots$, let $\hat{q}_k(t) = \mathbb{P}\left(\sum_{i=1}^{\infty} Z_i^{(k)}(t) > 0\right)$. Then, Windridge shows that there exists a constant $\hat{c} \in (0, 1]$ such that, for any $a < \min\{\tau, -r\}$,

$$\hat{q}_k(t) = \hat{c}ke^{rt} (1 + O(ke^{-at})) \quad \text{as } t \rightarrow \infty, \quad (7.5)$$

for any $k \geq 1$. The constant $\hat{c} = \lim_{t \rightarrow \infty} e^{-rt} \hat{q}_1(t)$. Note that for some practical purposes, $\hat{q}_k(t)$ may be of more interest than $q_k(t)$, since type-0 individuals are unable to transmit infection. In particular, in Appendix B we sketch the argument that for the case where $r > -\gamma$ an analogous result to (7.5) holds, i.e., for $k \geq 1$,

$$q_k(t) \sim cke^{rt} \quad \text{as } t \rightarrow \infty, \quad \text{where } c = \lim_{t \rightarrow \infty} e^{-rt} q_1(t) > 0. \quad (7.6)$$

(For real-valued functions, f and g say, $f(t) \sim g(t)$ as $t \rightarrow \infty$ if $\lim_{t \rightarrow \infty} f(t)/g(t) = 1$.)

For the case where $r < -\gamma$ (so, from (3.4), $\mu_{\tilde{D}} < 2$) we show in Appendix B that if $\mu_{\tilde{D}^2} < \infty$ then, for $k \geq 0$,

$$q_k(t) \sim (1 - k\mu_{\tilde{D}-2}^{-1})e^{-\gamma t} \quad \text{as } t \rightarrow \infty. \quad (7.7)$$

Note that in this case the asymptotic behaviour of the survival probability $q_k(t)$ is independent of the infection rate τ . The case $r < -\gamma$ could occur, for example, at the end of an epidemic where $\tau \gg \gamma$. Such an epidemic would consist primarily of transmission events at early times, with the late behaviour dominated by recovery events.

7.3 Late behaviour of the supercritical case

An approximation to $q_k(t)$ in the supercritical case ($r > 0$) can be obtained by exploiting the fact that a supercritical branching process conditioned on extinction is probabilistically equivalent to a subcritical branching process. For $k \in \mathbb{N}_0$, let $T^{(k)} = \inf\{t \geq 0 : Z^{(k)}(t) = 0\}$ denote the extinction time of the branching process given that it starts with one individual of type k , where $T^{(k)} = \infty$ if the branching process survives forever, and let $\pi_k = \mathbb{P}(T^{(k)} < \infty) = \pi_k(\infty)$ be the probability that the branching process ultimately goes extinct. Then,

$$q_k(t) = 1 - \pi_k + \pi_k \mathbb{P}(T^{(k)} > t | T^{(k)} < \infty). \quad (7.8)$$

Let $\{\tilde{\mathbf{Z}}^{(k)}(t) : t \geq 0\}$ be distributed as $\{\mathbf{Z}^{(k)}(t) : t \geq 0 | T^{(k)} < \infty\}$. Then it follows from Waugh [32], Section 5, that $\{\tilde{\mathbf{Z}}^{(k)}(t) : t \geq 0\}$ is also a continuous-time multitype Markov branching process, in which the lifetime of a typical type- k individual has an exponential distribution with rate $\gamma + \tau k$, as at (2.1), but when it dies its offspring is now distributed as follows:

$$\begin{aligned} \mathbb{P}(\text{Offspring} = \emptyset) &= \frac{1}{\pi_k} \frac{\gamma}{\gamma + \tau k}, \\ \mathbb{P}(\text{Offspring} = \{k-1, l\}) &= \frac{\pi_{k-1}\pi_l}{\pi_k} \frac{\tau k \tilde{d}_{l+1}}{\gamma + \tau k} \quad (l \in \mathbb{N}_0). \end{aligned} \quad (7.9)$$

Suppose now that there is a maximum degree size k_{\max} . Let $\tilde{\Omega} = [\tilde{\Omega}_{l,k}]$ be the $(k_{\max} + 1) \times (k_{\max} + 1)$ matrix with elements given by

$$\tilde{\Omega}_{l,k} = \frac{\pi_{l-1}\pi_k}{\pi_l} \tau l (\tilde{d}_{k+1} + \delta_{l,k+1}) - (\gamma + \tau l) \delta_{l,k} \quad (l, k \in \mathcal{K}).$$

Then, recalling (3.2),

$$\mathbb{E} \left[\tilde{Z}^{(k)}(t) \right] = \mathbf{u}_k^\top \mathbf{e}^{\tilde{\mathbf{\Omega}} t} \mathbf{1} , \quad (7.10)$$

where $\tilde{Z}^{(k)}(t) = \tilde{Z}_0^{(k)}(t) + \tilde{Z}_1^{(k)}(t) + \dots + \tilde{Z}_{k_{\max}}^{(k)}(t)$.

Let \tilde{r} denote the dominant eigenvalue of $\tilde{\mathbf{\Omega}}$ and note that $\tilde{r} < 0$. For $t \geq 0$ and $k = 0, 1, \dots, k_{\max}$, let $\tilde{q}_k(t) = \mathbb{P} \left(\tilde{Z}^{(k)}(t) > 0 \right)$ be the probability that the branching process $\{\tilde{\mathbf{Z}}^{(k)}(t) : t \geq 0\}$ is not extinct at time t given that it starts with one individual of type k . Then we expect that arguments similar to those used in the proof of Heinzmann [18], Theorem 3.1, will show that there exists constants $\tilde{c}_1, \tilde{c}_2, \dots, \tilde{c}_{k_{\max}}$, satisfying $0 < \tilde{c}_k < \infty$, such for $k = 1, 2, \dots, k_{\max}$,

$$\tilde{q}_k(t) = \tilde{c}_k e^{\tilde{r}t} (1 + o(e^{-\tilde{\gamma}t})) \quad \text{as } t \rightarrow \infty , \quad (7.11)$$

for any $\tilde{\gamma} > 0$. It then follows using (7.8) that

$$q_k(t) = 1 - \pi_k + \pi_k \tilde{c}_k e^{\tilde{r}t} (1 + o(e^{-\tilde{\gamma}t})) \quad \text{as } t \rightarrow \infty . \quad (7.12)$$

Heinzmann [18], Theorem 3.1, cannot be applied directly as it assumes that the matrix $\tilde{\mathbf{\Omega}}$ is irreducible. We do not consider it here but we expect that Heinzmann's proof can be extended to our situation. If we assume that type-0 individuals are dead and only consider initial individuals of types $1, 2, \dots, k_{\max} - 1$ (recall that only initial infectives can have type k_{\max}) then $\tilde{\mathbf{\Omega}}$ becomes a $(k_{\max} - 1) \times (k_{\max} - 1)$ irreducible matrix. Heinzmann [18], Theorem 3.1, then yields (7.11); note that now π_k is replaced by $\bar{\pi}_k = 1 - \lim_{t \rightarrow \infty} \bar{q}_k(t)$ ($k = 1, 2, k_{\max} - 1$) in the definition of $\tilde{\mathbf{\Omega}}$ and $\tilde{q}_k(t) = \bar{q}_k(t) = \mathbb{P} \left(\sum_{i=1}^{k_{\max}-1} \tilde{Z}_i^{(k)}(t) > 0 \right)$. The approximation (7.12) then holds with $q_k(t)$ and π_k replaced by $\bar{q}_k(t)$ and $\bar{\pi}_k$, respectively. Moreover, if we then let $\tilde{\mathbf{f}}_1^\top$ and $\tilde{\mathbf{b}}_1$ be left and right eigenvectors of $\tilde{\mathbf{\Omega}}$ corresponding to the eigenvalue \tilde{r} , satisfying $\tilde{\mathbf{f}}_1^\top \tilde{\mathbf{b}}_1 = 1$, then $\tilde{c}_k = \left(\mathbf{u}_k^\top \tilde{\mathbf{b}}_1 \right) h^*$, where $h^* = \lim_{t \rightarrow \infty} e^{-\tilde{r}t} \tilde{\mathbf{f}}_1^\top \tilde{\mathbf{q}}(t)$ and $\tilde{\mathbf{q}}(t) = (\tilde{q}_1(t), \tilde{q}_2(t), \dots, \tilde{q}_{k_{\max}-1}(t))^\top$. Unfortunately, unlike with $\mathbf{\Omega}$, there does not appear to be closed-form expressions for \tilde{r} and its associated eigenvectors.

8 Fluctuations in the emerging phase of a major outbreak

We now consider the early behaviour of supercritical epidemics that take off (i.e. do not go extinct early on but ultimately end owing to long-term depletion of susceptibles). The early stages of such an epidemic are approximated by the branching process defined in Section 2.3 but conditioned on non-extinction. It is straightforward to adapt the results on means and variances in Sections 3 and 4 to condition on $Z^{(k)}(t) > 0$. Elementary calculation shows that, for $t \geq 0$ and $k \in \mathbb{N}_0$,

$$\begin{aligned} \mathbb{E} \left[Z^{(k)}(t) \middle| Z^{(k)}(t) > 0 \right] &= \frac{\mathbb{E} [Z^{(k)}(t)]}{q_k(t)} , \\ \text{var} \left(Z^{(k)}(t) \middle| Z^{(k)}(t) > 0 \right) &= \frac{\text{var} (Z^{(k)}(t))}{q_k(t)} - \pi_k(t) \left(\frac{\mathbb{E} [Z^{(k)}(t)]}{q_k(t)} \right)^2 . \end{aligned}$$

Expressions for $\mathbb{E} \left[Z^{(k)}(t) \middle| Z^{(k)}(t) > 0 \right]$ and $\text{var} \left(Z^{(k)}(t) \middle| Z^{(k)}(t) > 0 \right)$ then follow using (3.15) and (4.10), respectively, though there is no closed-form formula for $q_k(t)$ or $\pi_k(t)$. Note

that, assuming $r > 0$ so $\pi_k < 1$,

$$\lim_{t \rightarrow \infty} \frac{\text{var} \left(Z^{(k)}(t) \middle| Z^{(k)}(t) > 0 \right)}{\mathbb{E} \left[Z^{(k)}(t) \middle| Z^{(k)}(t) > 0 \right]^2} = (1 - \pi_k) \lim_{t \rightarrow \infty} \frac{\text{var} \left(Z^{(k)}(t) \right)}{\mathbb{E} \left[Z^{(k)}(t) \right]^2} - \pi_k ,$$

which depends on the degree k of the initial infective.

The diffusion approximation studied in Graham and House [16] corresponds to the case where the number of infectives at time $t = 0$ is large. Return to the case where there is a maximum degree k_{\max} and suppose that the branching process does not go extinct. Then it follows from Athreya and Ney [1], Theorem V.7.2, that, for any $k \in \mathcal{K}$,

$$\frac{\mathbf{Z}^{(k)}(t)}{Z^{(k)}(t)} \xrightarrow{\text{a.s.}} \mathbf{w}_1 \quad \text{as } t \rightarrow \infty ,$$

where \mathbf{w}_1 , given by (3.10), is a left eigenvector of $\mathbf{\Omega}$ corresponding to the dominant eigenvalue $\lambda_1 = r$. Thus if an epidemic takes off and is still in its exponentially growing phase then the relative frequencies of the different types of infectives will be close to \mathbf{w}_1 . Hence, we now assume that the initial number of individuals in the branching process $Z(0) = K$, where K is large, and that $Z_i(0) \approx w_{1,i}K$ for $i \in \mathcal{K}$. Label the initial individuals $1, 2, \dots, K$. Then, for $t \geq 0$, the total population size is $Z(t) = \hat{Z}_1(t) + \hat{Z}_2(t) + \dots + \hat{Z}_K(t)$, where $\hat{Z}_i(t)$ denotes the total number of descendants of the initial individual i that are alive at time t , including i itself if it is still alive. Thus, $\mathbb{E}[Z(t)] = \sum_{i=1}^K \mathbb{E}[\hat{Z}_i(t)]$, for all $t \geq 0$, and, since the processes $\{\hat{Z}_i(t) : t \geq 0\}$ ($i = 1, 2, \dots, K$) are mutually independent, $\text{var}(Z(t)) = \sum_{i=1}^K \text{var}(\hat{Z}_i(t))$, for all $t \geq 0$, and $\text{cov}(Z(t), Z(s)) = \sum_{i=1}^K \text{cov}(\hat{Z}_i(t), \hat{Z}_i(s))$, for all $t, s \geq 0$.

Note that (3.9) implies that

$$\mathbf{w}_1^\top \mathbf{n} = \sum_{i=0}^{k_{\max}} i w_{1,i} = \mu_{\tilde{D}-2} \quad \text{and} \quad \mathbf{w}_1^\top \mathbf{n}_2 = \sum_{i=0}^{k_{\max}} i^2 w_{1,i} = \frac{\mu_{\tilde{D}-2} \mu_{(\tilde{D}-1)^2+1}}{\mu_{\tilde{D}}} .$$

Assuming that the above approximation is exact, then, for $t \geq 0$, it follows from (3.15) that

$$\mathbb{E}[Z(t)] = K \mathbf{w}_1^\top \mathbf{m}(t) = K e^{rt} \quad (8.1)$$

and, after a little algebra, it follows from (4.10) that

$$\begin{aligned} \text{var}(Z(t)) &= K \mathbf{w}_1^\top \mathbf{v}(t) \\ &= K \left(\gamma \left[I_9(t) + \mu_{\tilde{D}}^{-1} \mu_{\tilde{D}-2}^{-1} \left(\sigma_{\tilde{D}}^2 + 2 \right) I_4(t) \right] + \tau \mu_{\tilde{D}-2}^{-1} \left[\mu_{\tilde{D}-2}^2 I_9(t) + \sigma_{\tilde{D}}^2 I_4(t) \right] \right) , \end{aligned} \quad (8.2)$$

where $\sigma_{\tilde{D}}^2 = \text{var}(\tilde{D})$ and

$$I_9(t) = \frac{e^{rt} (e^{rt} - 1)}{r} .$$

Comparison of (8.1) and (8.2) with the diffusion-based result of Graham and House [16] in the limit of large t gives agreement when $\gamma = 0$ (i.e. for the SI model) but not for $\gamma > 0$. We believe that this is due to the fact that the diffusion model was only four dimensional, so a heuristic argument (given in Section 3.3 of Graham and House [16], which gave results that

were in good agreement with simulation) about the neighbourhood of an infective node had to be made, in contrast to the approach here that deals with each effective degree explicitly and so has $k_{\max} + 1$ dimensions. The argument about the neighbourhood around an infective tries to account for correlations caused by variability in recovery times, and so if $\gamma \rightarrow 0$ then these correlations do not exist.

Recent work by Constable and McKane [8] considered the reduction of high-dimensional stochastic models to low-dimensional diffusions and this approach was shown to be asymptotically exact for some systems in the small-noise limit by Parsons and Rogers [29]. It is an open question whether the argument in Section 3.3 of Graham and House [16] could be justified rigorously by a similar argument, however we note that a branching-process approach makes fewer assumptions than a low-dimensional diffusion limit and so will be more generally applicable.

Considering further results that can be obtained, it follows from (5.7) and a little algebra that, for $0 \leq t \leq s$,

$$\begin{aligned} \text{cov}(Z(t), Z(s)) &= e^{-\gamma(s-t)} \text{var}(Z(t)) + K\mu_{\bar{D}-2}^{-1} \left(e^{\gamma(s-t)} - e^{-\gamma(s-t)} \right) \\ &\quad \times \left\{ \gamma\mu_{\bar{D}}^{-1} \left[\mu_{(\bar{D}-1)^2+1} I_9(t) - \left(\sigma_{\bar{D}}^2 + 2 \right) I_6(t) \right] \right. \\ &\quad \left. + \tau \left[\mu_{\bar{D}-2}^2 I_9(t) - \sigma_{\bar{D}}^2 I_6(t) \right] \right\} . \end{aligned} \quad (8.3)$$

It seems plausible that these results extend to the case when there is no maximal degree but that would involve results for countably infinite matrices which we do not consider here.

Recall that the processes $\{\hat{Z}_i(t) : t \geq 0\}$ ($i = 1, 2, \dots, K$) are mutually independent. It follows using the central limit theorem that, for sufficiently large K , the process $\{Z(t) : t \geq 0\}$, which approximates the prevalence of infection during the early growth of an epidemic, is approximately Gaussian with mean function given by (8.1) and covariance function given by (8.3).

9 Numerical examples

9.1 Forward simulations

We conducted a series of numerical experiments to provide specific examples of the general results presented here. $M = 10^4$ Monte Carlo simulations were performed on three different configuration model networks, each of size $N = 10^4$, and with the degree distributions shown in Figure 1 Row (i) that have the same mean but different higher moments. The first of these, $D^{(1)}$, has $d_3^{(1)} = 1$, the second, $D^{(2)}$, has $d_1^{(2)} = d_2^{(2)} = d_3^{(2)} = d_4^{(2)} = d_5^{(2)} = 1/5$, and the third, $D^{(3)}$, has $d_1^{(3)} = 1/8$, $d_3^{(3)} = 5/6$, $d_9^{(3)} = 1/24$. (Note that each Monte Carlo simulation consisted of first simulating a network and then simulating a single epidemic on it.) Two different scenarios were considered. In the first – most commonly considered in the literature when simulations are compared to analytic approaches – time = 0 was defined as the first point when prevalence is at a given level, K . In our simulations we took $K = 100$, but in general K should take a value where the probability of extinction has become negligible, but the depletion of the susceptible population has not had a significant effect on the epidemic dynamics. In the second, each epidemic was started from one node, picked uniformly at random, so the probability of extinction played a major role. This scenario is less commonly considered when comparing real-time simulated epidemics to

differential equation models because the latter are typically designed to hold when the epidemic is already established.

Since analytic results for the probabilities of extinction $\pi_k(t)$ are not available, the branching process results required numerical integration of ordinary differential equations (in our case using Runge-Kutta methods). We stress that the computational effort required to do this is much less than that involved in performing Monte Carlo simulations, and has the benefit of not depending on N .

The results for the first approach (restarting time at the first time prevalence reaches 100) are given in Figure 1. Row (ii) shows sample trajectories (which all agree on prevalence at time 0). Row (iii) shows the simulated mean after time 0 on a logarithmic scale, which initially grows at the constant rate predicted by the branching process model, and then reduces as the susceptible population is depleted. Row (iv) shows the variance, which has not converged to its asymptotic growth regime by the time prevalence is equal to 100, an effect that is captured by the branching process model. The variance does not take its largest value at the peak prevalence, but instead has local maxima before and after the peak.

Figure 2 shows the results for the second approach in which there is one randomly chosen initial infective at time 0. Row (i) shows some sample trajectories. Row (ii) shows the extinction probabilities, which are accurately captured in the branching process model until very close to the end of the epidemic when prevalence is low and extinction becomes likely again. Row (iii) shows the mean, which does not start growing at a constant rate with the convergence rate accurately captured in the branching process model. Row (iv) shows the variance and convergence onto its asymptotic value; in this case there is a single maximum just before the peak in prevalence.

9.2 Statistical inference

In order to demonstrate the potential use of the real-time effective degree branching process model for statistical inference, we carried out a simulation study. Here we simulated one epidemic that took off on a configuration model network of size $N = 10^6$ with degree distribution $D^{(3)}$ as in the right-hand column of Figure 1 ($d_1^{(3)} = 1/8$, $d_3^{(3)} = 5/6$, $d_9^{(3)} = 1/24$) and true rates $\tau_0 = 2$, $\gamma_0 = 1$. Letting $I(t)$ be the prevalence of infection in the network model, we set time $t = 0$ when $I(t) = 100$ for the first time and make 40 evenly-spaced observations (with gap $\delta t = 0.05$ between each) of $I(t)$ up to $t_{\text{end}} = 2$.

We then define an approximate likelihood based on the methods of Ross *et al.* [30], in which a Gaussian process approximation based on known first and second moments is used, which will be more accurate for larger N , larger $I(0)$ and smaller δt . There should, however, be no *a priori* obstacle to fitting our model to data on smaller populations even with incomplete data, for example by using Markov chain Monte Carlo methods to perform multiple imputation as suggested by O'Neill and Roberts [28].

Explicitly, we let the probability density function f for sequential observations be given by

$$f(I(t + \delta t)|I(t)) = \mathcal{N}(\mathbb{E}[Z(t + \delta t)|Z(t) = I(t)], \text{var}(Z(t + \delta t)|Z(t) = I(t))) , \quad (9.1)$$

where $\mathcal{N}(m, V)$ is the probability density function of a normal distribution with mean m and variance V , and the expectation and variance of $Z(t + \delta t)$ are given by the results of Section 8 above. The likelihood is then

$$L = \prod_{t \in \{0, \delta t, \dots, t_{\text{end}} - \delta t\}} f(I(t + \delta t)|I(t)) . \quad (9.2)$$

We consider values of this likelihood across the range of rate constant parameters τ and γ under two different degree distributions: the correct one, $D^{(3)}$, and a misspecified degree distribution $D^{(1)}$, which is the one used in the left-hand column of Figure 1 ($d_3^{(1)} = 1$).

Figure 3 shows the first quarter of the simulated epidemic together with the Gaussian process approximation, as well as likelihood surfaces for the correct and misspecified degree distributions. Performing maximum likelihood estimation using MATLAB's `mle()` function allows us to obtain point estimates for parameters $\hat{\tau}$ and $\hat{\gamma}$, as well as asymptotic 95% confidence intervals. We also use MATLAB's `mlecov()` function to obtain the parameter covariance matrix \hat{C} from the inverse Hessian. We quote results to 2 significant figures; the asymptotic approximations also give very slightly negative lower confidence intervals for $\hat{\gamma}$ which we round up to 0. For the correct degree distribution we obtain

$$\begin{pmatrix} \hat{\tau}^{(3)} \\ \hat{\gamma}^{(3)} \end{pmatrix} = \begin{pmatrix} 1.9 & [1.4, 2.4] \\ 0.8 & [0, 1.7] \end{pmatrix}, \quad \hat{C}^{(3)} = \begin{pmatrix} 0.069 & 0.11 \\ 0.11 & 0.19 \end{pmatrix}, \quad (9.3)$$

and for the misspecified degree distribution we obtain

$$\begin{pmatrix} \hat{\tau}^{(1)} \\ \hat{\gamma}^{(1)} \end{pmatrix} = \begin{pmatrix} 3.2 & [2.3, 4.0] \\ 0.8 & [0, 1.7] \end{pmatrix}, \quad \hat{C}^{(1)} = \begin{pmatrix} 0.20 & 0.20 \\ 0.20 & 0.20 \end{pmatrix}. \quad (9.4)$$

This shows that knowledge of the correct distribution allows both τ and γ to be estimated; although as would be expected the early asymptotic growth rate r is much more closely constrained by simulated data than other directions in parameter space. It also shows that misspecification of the degree distribution allows r to be identified, but biases the estimate of, in this case, τ .

10 Concluding comments

10.1 Summary of results

In this paper, we have provided explicit closed-form expressions for the real-time mean, variance and covariance function for disease prevalence of the Markovian SIR model on a configuration model network, as well as deriving differential equations for the probabilities of extinction over time that are relatively numerically cheap to solve. These allow for a more explicit treatment of e.g. rates of convergence to asymptotic behaviour than has previously been possible.

10.2 Future directions

We believe that the methods of real-time, multi-type branching processes could be more widely applied in infectious disease epidemiology, since they provide results concerning extinction and variance that are not available using deterministic differential equation models. For example, the effective-degree based methodology presented here may be extended to include degree correlation (e.g. in the sense of Newman [26]) by keeping track of the actual, as well as effective, degrees of individuals, though the type space becomes larger and explicit analytic results are unlikely to be available. We note that there is increasing interest in the eradication of infections (e.g. Klepac *et al.* [22]) and that arguably calculating extinction probabilities and variability in outbreak sizes is of equal or greater importance in this context than calculation of mean behaviour.

The explicit closed-form expressions derived have the potential to enhance statistical work on epidemic prevalence curves. In particular, many empirically observed epidemics

of human pathogens exhibit more variability around the trend than simple models would predict (see Black *et al.* [6], particularly Section 1, for a discussion of this), which can bias parameter estimation if an insufficiently variable model is used. Application of our methods to real data would be an interesting extension of our work.

The possibility of a more general non-Markovian stochastic epidemic being approximated by an appropriate real-time branching process is raised by the results of Barbour and Reinert [5] and it would be interesting to investigate whether our analysis could be adapted to this scenario.

Finally, there is the question of low-dimensional PGF-based modelling of the whole network epidemic that incorporates stochasticity accurately. For example, the work of Miller [24] considered accounting for early fluctuations and Graham and House [16] considered a diffusion approximation once early fluctuations were negligible, but the results presented here as well as those of Barbour and Reinert [5] suggest that a more unified low-dimensional stochastic approach that explicitly models early fluctuations may be possible.

A Eigenvalues of Ω

Let $\mathbf{A} = [a_{l,k}] = \Omega - \lambda \mathbf{I}$. Observe that $a_{0,0} = -(\lambda + \gamma)$, $a_{0,k} = 0$ for $k = 1, 2, \dots, k_{\max}$, $a_{l,k_{\max}} = 0$ for $l = 1, 2, \dots, k_{\max} - 1$ and $a_{k_{\max},k_{\max}} = -(k_{\max}\tau + \lambda + \gamma)$. Thus, expanding $|\mathbf{A}|$ along the 0-th row and then the cofactor $A_{0,0}$ down the last column yields

$$|\mathbf{A}| = (\lambda + \gamma)(k_{\max}\tau + \lambda + \gamma)|\mathbf{B}|, \quad (\text{A.1})$$

where

$$\mathbf{B} = \begin{bmatrix} \tau(\tilde{d}_2 - 1) - \lambda - \gamma & \tau\tilde{d}_3 & \cdots & \tau\tilde{d}_{k_{\max}} \\ 2\tau(\tilde{d}_2 + 1) & 2\tau(\tilde{d}_3 - 1) - \lambda - \gamma & \cdots & 2\tau\tilde{d}_{k_{\max}} \\ \vdots & \vdots & \ddots & \vdots \\ (k_{\max} - 1)\tau\tilde{d}_2 & (k_{\max} - 1)\tau\tilde{d}_3 & \cdots & (k_{\max} - 1)\tau(\tilde{d}_{k_{\max}} - 1) - \lambda - \gamma \end{bmatrix}.$$

More precisely, \mathbf{B} is the $(k_{\max} - 1) \times (k_{\max} - 1)$ matrix with elements

$$b_{l,k} = \tau l \left(\tilde{d}_{k+1} + \delta_{l,k+1} \right) - (\gamma + \tau l + \lambda) \delta_{l,k} \quad (l, k = 1, 2, \dots, k_{\max} - 1).$$

Subtracting $l \times$ the first row of \mathbf{B} from the l -th row of \mathbf{B} , for $l = 2, 3, \dots, k_{\max} - 1$, now gives $|\mathbf{B}| = |\mathbf{C}|$, where

$$\mathbf{C} = \begin{bmatrix} \tau(\tilde{d}_2 - 1) - \lambda - \gamma & \tau\tilde{d}_3 & \cdots & \tau\tilde{d}_{k_{\max}} \\ 2(2\tau + \lambda + \gamma) & -2\tau - \lambda - \gamma & \cdots & 0 \\ 3(\tau + \lambda + \gamma) & 3\tau & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ (k_{\max} - 1)(\tau + \lambda + \gamma) & 0 & \cdots & -(k_{\max} - 1)\tau - \lambda - \gamma \end{bmatrix}$$

has elements given by

$$\begin{aligned} c_{1,k} &= \tau\tilde{d}_{k+1} - (\gamma + \tau + \lambda)\delta_{1,k} & (k = 1, 2, \dots, k_{\max} - 1), \\ c_{l,k} &= \tau l \delta_{l,k+1} - (\gamma + \tau l + \lambda)\delta_{l,k} + l(\gamma + \tau + \lambda)\delta_{1,k} & (l, k = 1, 2, \dots, k_{\max} - 1). \end{aligned}$$

In particular, $c_{1,1} = \tau(\tilde{d}_2 - 1) - \gamma - \lambda$, $c_{1,k} = \tau\tilde{d}_{k+1}$ for $k = 2, 3, \dots, k_{\max} - 1$, and $c_{l,k} = 0$ for $2 \leq l < k \leq k_{\max} - 1$. Thus, adding $k \times$ the k -th column of \mathbf{C} to the first column of \mathbf{C} , for $k = 2, 3, \dots, k_{\max} - 1$, yields $|\mathbf{C}| = |\mathbf{D}|$, where

$$\mathbf{D} = \begin{bmatrix} \tau((\sum_{l=0}^{k_{\max}} l\tilde{d}_{l+1}) - 1) - \lambda - \gamma & \tau\tilde{d}_3 & \cdots & \tau\tilde{d}_{k_{\max}} \\ 0 & -2\tau - \lambda - \gamma & \cdots & 0 \\ 0 & 3\tau & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -(k_{\max} - 1)\tau - \lambda - \gamma \end{bmatrix}.$$

Note that $d_{1,1} = \tau \left[\left(\sum_{l=0}^{k_{\max}} l\tilde{d}_{l+1} \right) - 1 \right] - \gamma - \lambda$ and, for $l \geq 2$, that $d_{l,1} = 0$, $d_{l,l} = -(\gamma + \tau + \lambda)$ and $d_{l,k} = 0$ for $k > l$. Thus expanding $|\mathbf{D}|$ down the first column gives

$$|\mathbf{D}| = \left\{ \tau \left[\left(\sum_{l=0}^{k_{\max}} l\tilde{d}_{l+1} \right) - 1 \right] - \gamma - \lambda \right\} (-1)^{k_{\max}-2} \prod_{l=2}^{k_{\max}-1} (l\tau + \gamma + \lambda).$$

Recalling (A.1) and $|\mathbf{B}| = |\mathbf{C}| = |\mathbf{D}|$, it follows that the eigenvalues of $\mathbf{\Omega}$ are given by (3.3).

B Late behaviour of subcritical survival probabilities

To bound the late probabilities of survival in the subcritical case, first note that due to the inability of type-0 individuals to transmit we have

$$q_0(t) = e^{-\gamma t}. \quad (\text{B.1})$$

Recalling that $q_k(t) = 1 - \pi_k(t)$, it follows from (7.2), with $k_{\max} = \infty$, that for $k > 0$,

$$\frac{dq_k}{dt} = -(\gamma + \tau k)q_k + \tau k q_{k-1} + \tau k (1 - q_{k-1}) \sum_{l=0}^{\infty} \tilde{d}_{l+1} q_l. \quad (\text{B.2})$$

This leads to the equation for $k = 1$ in the form

$$\frac{dq_1}{dt} = r q_1 + h(t), \quad (\text{B.3})$$

where

$$h(t) = \underbrace{\tau e^{-\gamma t} \left(1 + \tilde{d}_1 - \sum_{k=0}^{\infty} \tilde{d}_{k+1} q_k(t) \right)}_{h_1(t)} + \underbrace{\tau \sum_{k=1}^{\infty} \tilde{d}_{k+1} (q_k(t) - k q_1(t))}_{h_2(t)}. \quad (\text{B.4})$$

We assume first that $r > -\gamma$. Integrating (B.3) gives

$$e^{-rt} q_1(t) = 1 - \int_0^t e^{-ru} h(u) du.$$

The limit $\lim_{t \rightarrow \infty} e^{-rt} q_1(t)$ therefore exists if r is within the region of convergence of the Laplace transform of h . Considering $h_1(t)$ in (B.4), since $q_k(t) \in [0, 1]$, for all k , we have that $\tilde{d}_1 \leq h_1(t) \leq \tau(1 + \tilde{d}_1)e^{-\gamma t}$, so the Laplace transform of h_1 converges by the assumption that $r > -\gamma$.

Now considering h_2 , we follow Windridge [33] and consider an initial individual with effective degree k , and stubs labelled by integer $i = 1, 2, \dots, k$. Let T_i be the time that the

individual and its progeny through stub i exist, and let T be the lifetime of the branching process. Then, using the Bonferroni inequalities as in [33], for $k \geq 1$,

$$\{T > t\} = \bigcup_{i=1}^k \{T_i > t\} , \quad q_k(t) \leq kq_1(t) \quad \text{and} \quad q_k(t) \geq kq_1(t) - k^2\mathbb{P}(T_1 > t, T_2 > t) . \quad (\text{B.5})$$

Therefore, we have that

$$0 \leq h_2(t) \leq \tau\mu_{(\tilde{D}-1)^2}\mathbb{P}(T_1 > t, T_2 > t) .$$

Now, from (7.4) we have that for some constant κ ,

$$q_k(t) \leq k\kappa e^{rt} . \quad (\text{B.6})$$

Writing R for the lifetime of the initial infective individual we have that

$$\mathbb{P}(T_1 > t, T_2 > t) = \mathbb{P}(R > t) + \mathbb{P}(T_1 > t, T_2 > t, R \leq t) .$$

We then recall that $\mathbb{P}(R > t) = q_0(t) = e^{-\gamma t}$, after which the argument follows closely that of [33] (as in the derivation of (B.13) below) and we find that

$$\lim_{t \rightarrow \infty} \int_0^\infty e^{-rt} \mathbb{P}(T_1 > t, T_2 > t) dt < \infty . \quad (\text{B.7})$$

Thus the Laplace transform of h_2 converges at r , whence $\lim_{t \rightarrow \infty} e^{-rt} q_1(t)$ is equal to a finite constant, c say. Note that c is strictly positive since $c \geq \lim_{t \rightarrow \infty} e^{-rt} \hat{q}_1(t) = \hat{c} > 0$. Further, (B.7) implies that $\lim_{t \rightarrow \infty} e^{-rt} \mathbb{P}(T_1 > t, T_2 > t) = 0$, and it follows from the two inequalities in (B.5) that, $\lim_{t \rightarrow \infty} e^{-rt} q_k(t) = kc$, for $k \geq 0$, proving (7.6).

We consider now the case when $r < -\gamma$. For $t \geq 0$ and $k = 0, 1, \dots$. Write

$$q_k(t) = e^{-\gamma t} + \tilde{q}_k(t) \quad \text{and} \quad \tilde{u}_k(t) = e^{\gamma t} \tilde{q}_k(t) , \quad (\text{B.8})$$

so $\tilde{q}_k(t)$ is the probability that the branching process has survived to time t but the initial individual has not, and $\tilde{u}_0(t) = 0$ for all $t \geq 0$. It follows using (B.2) that

$$\frac{d\tilde{u}_1}{dt} = \tau \left[-\tilde{u}_1 + (1 - e^{-\gamma t}) \left(1 + \sum_{l=1}^{\infty} \tilde{d}_{l+1} \tilde{u}_l \right) \right] . \quad (\text{B.9})$$

The Bonferroni inequalities yield that, for $k \geq 1$,

$$k\tilde{q}_1(t) - k^2\mathbb{P}(R < t, T_1 > t, T_2 > t) \leq \tilde{q}_k(t) \leq k\tilde{q}_1(t) , \quad (\text{B.10})$$

so

$$0 \leq \tilde{u}_k(t) \leq k\tilde{u}_1(t) , \quad (\text{B.11})$$

and (B.9) implies that

$$\frac{d\tilde{u}_1}{dt} \leq \tau (1 + \mu_{\tilde{D}-2} \tilde{u}_1) ,$$

whence, for all $t \geq 0$, recalling that $\mu_{\tilde{D}-2} < 0$,

$$0 \leq \tilde{u}_1(t) \leq -\frac{1}{\mu_{\tilde{D}-2}} (1 - e^{\tau\mu_{\tilde{D}-2}t}) \leq -\frac{1}{\mu_{\tilde{D}-2}} . \quad (\text{B.12})$$

Conditioning on the lifetime of the initial individual in the branching process,

$$\begin{aligned}\mathbb{P}(R \leq t, T_1 > t, T_2 > t) &= \int_{u=0}^t \gamma e^{-\gamma u} \left[\int_{v=0}^u \tau e^{-\tau v} q_1(t-v) dv \right]^2 du \\ &\leq \int_{u=0}^t \gamma e^{-\gamma u} \left[\int_{v=0}^u \tau e^{-\tau v} \left(-\frac{\mu_{\tilde{D}-1}}{\mu_{\tilde{D}-2}} \right) e^{-\gamma(t-v)} dv \right]^2 du ,\end{aligned}$$

since (B.8) and (B.12) imply that $q_1(t) \leq -\frac{\mu_{\tilde{D}-1}}{\mu_{\tilde{D}-2}} e^{-\gamma t}$. Elementary integration then shows that there exists $c' = c'(\mu, \tau, \mu_{\tilde{D}}) < \infty$ such that, for all $t \geq 0$,

$$e^{\gamma t} \mathbb{P}(R \leq t, T_1 > t, T_2 > t) \leq \begin{cases} c' e^{-\min\{\gamma, 2\tau\}t} & \text{if } \gamma \neq 2\tau , \\ c' t e^{-\gamma t} & \text{if } \gamma = 2\tau . \end{cases} \quad (\text{B.13})$$

The differential equation (B.9) may be written in the form

$$\frac{d\tilde{u}_1}{dt} = \tau(1 + \tilde{u}_1) - \underbrace{\tau e^{-\gamma t} \left(1 + \sum_{l=1}^{\infty} \tilde{d}_{l+1} \tilde{u}_l \right)}_{h_3(t)} - \underbrace{\tau \sum_{l=2}^{\infty} \tilde{d}_{l+1} (l\tilde{u}_1 - \tilde{u}_l)}_{h_4(t)} ,$$

whence

$$\tilde{u}_1(t) = -\frac{1}{\mu_{\tilde{D}-2}} (1 - e^{\tau\mu_{\tilde{D}-2}t}) - \tau e^{\tau\mu_{\tilde{D}-2}t} \int_0^t e^{-\tau\mu_{\tilde{D}-2}u} (h_3(u) + h_4(u)) du . \quad (\text{B.14})$$

Now (B.11) and (B.12) imply that $0 \leq h_3(t) \leq -\frac{1}{\mu_{\tilde{D}-2}} e^{-\gamma t}$, whence

$$\lim_{t \rightarrow \infty} e^{\tau\mu_{\tilde{D}-2}t} \int_0^t e^{-\tau\mu_{\tilde{D}-2}u} h_3(u) du = 0 . \quad (\text{B.15})$$

Further, it follows using (B.10) and (B.13) that $0 \leq h_4(t) \leq \mu_{(\tilde{D}-1)2} c' e^{-\min\{\gamma, 2\tau\}t}$, where $e^{-\min\{\gamma, 2\tau\}t}$ is replaced by $te^{-\gamma t}$ if $\gamma = 2\tau$, whence (B.15) also holds when $h_3(u)$ is replaced by $h_4(u)$. Letting $t \rightarrow \infty$ in (B.14) yields (7.7).

Acknowledgements

We gratefully acknowledge support from the Isaac Newton Institute for Mathematical Sciences, Cambridge, where we held Visiting Fellowships under the Infectious Disease Dynamics programme and its follow-up meeting, during which this work was initiated. TH is supported by the Engineering and Physical Sciences Research Council. We would like to thank Josh Ross for helpful comments on this manuscript.

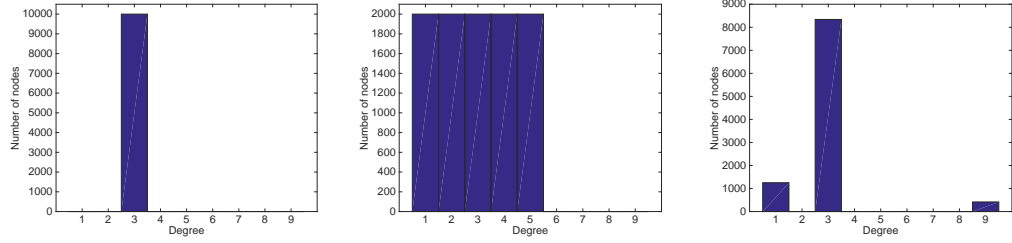
References

- [1] Athreya, K.B., Ney, P.E.: Branching Processes. Springer-Verlag, Berlin (1972)
- [2] Bailey, N.T.J.: The Mathematical Theory of Epidemics. Griffin, London (1957)
- [3] Ball, F., Donnelly, P.: Strong approximations for epidemic models. Stochastic Processes and their Applications **55**(1), 1–21 (1995)

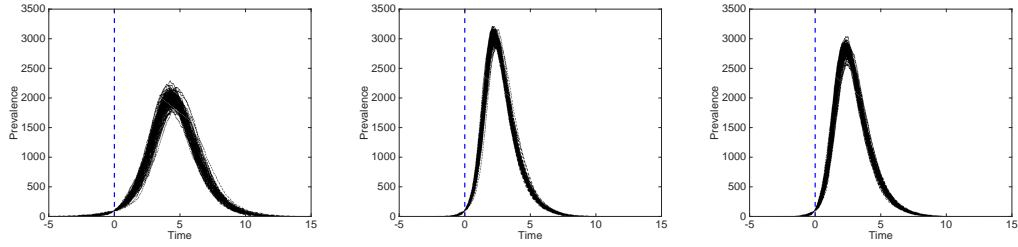
- [4] Ball, F., Neal, P.: Network epidemic models with two levels of mixing. *Mathematical Biosciences* **212**(1), 69–87 (2008)
- [5] Barbour, A., Reinert, G.: Approximating the epidemic curve. *Electronic Journal of Probability* **18**(54), 1–30 (2013)
- [6] Black, A.J., House, T., Keeling, M.J., Ross, J.V.: The effect of clumped population structure on the variability of spreading dynamics. *Journal of Theoretical Biology* **359**, 45–53 (2014)
- [7] Bohman, T., Piccollelli, M.: SIR epidemics on random graphs with a fixed degree sequence. *Random Structures and Algorithms* **41**(2), 179–214 (2012)
- [8] Constable, G.W.A., McKane, A.J.: Fast-mode elimination in stochastic metapopulation models. *Physical Review E* **89**(3), 032,141 (2014)
- [9] Daley, D.J., Vere-Jones, D.: *An Introduction to the Theory of Point Processes. Probability and Its Applications*. Springer, New York (1988)
- [10] Danon, L., Ford, A.P., House, T., Jewell, C.P., Keeling, M.J., Roberts, G.O., Ross, J.V., Vernon, M.C.: Networks and the epidemiology of infectious disease. *Interdisciplinary Perspectives on Infectious Diseases* **2011**, 1–28 (2011)
- [11] Decreusefond, L., Dhersin, J.S., Moyal, P., Tran, V.C.: Large graph limit for an SIR process in random network with heterogeneous connectivity. *The Annals of Applied Probability* **22**(2), 541–575 (2012)
- [12] Dorman, K., Sinsheimer, J., Lange, K.: In the garden of branching processes. *SIAM Review* **46**(2), 202–229 (2004)
- [13] Durrett, R.: *Random Graph Dynamics*. Cambridge University Press (2007)
- [14] Eames, K.T.D., Keeling, M.J.: Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *PNAS* **99**(20), 13,330–13,335 (2002)
- [15] Ethier, S.N., Kurtz, T.G.: *Markov processes: characterization and convergence*. Wiley series in probability and mathematical statistics. John Wiley and Sons, Hoboken, New Jersey (1986)
- [16] Graham, M., House, T.: Dynamics of stochastic epidemics on heterogeneous networks. *Journal of Mathematical Biology* **68**(7), 1583–1605 (2014)
- [17] Heesterbeek, H., Anderson, R.M., Andreasen, V., Bansal, S., De Angelis, D., Dye, C., Eames, K.T.D., Edmunds, W.J., Frost, S.D.W., Funk, S., Hollingsworth, T.D., House, T., Isham, V., Klepac, P., Lessler, J., Lloyd-Smith, J.O., Metcalf, C.J.E., Mollison, D., Pellis, L., Pulliam, J.R.C., Roberts, M.G., Viboud, C., Isaac Newton Institute IDD Collaboration: Modeling infectious disease dynamics in the complex landscape of global health. *Science* **347**(6227) (2015)
- [18] Heinzmann, D.: Extinction times in multitype Markov branching processes. *Journal of Applied Probability* **46**(1), 296–307 (2009)
- [19] Holme, P.: Extinction times of epidemic outbreaks in networks. *PLoS ONE* **8**(12), e84,429 (2013)

- [20] House, T., Keeling, M.J.: Insights from unifying modern approximations to infections on networks. *Journal of The Royal Society Interface* **8**(54), 67–73 (2010)
- [21] Janson, S., Luczak, M., Windridge, P.: Law of large numbers for the SIR epidemic on a random graph with given degrees. *Random Structures and Algorithms* **45**(4), 726–763 (2014)
- [22] Klepac, P., Metcalf, C.J.E., McLean, A.R., Hampson, K.: Towards the endgame and beyond: complexities and challenges for the elimination of infectious diseases. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **368**(1623) (2013)
- [23] Miller, J.C.: A note on a paper by Erik Volz: SIR dynamics in random networks. *Journal of Mathematical Biology* **62**(3), 349–358 (2011)
- [24] Miller, J.C.: Epidemics on networks with large initial conditions or changing structure. *PLoS ONE* **9**(7), e101,421 (2014)
- [25] Molloy, M., Reed, B.: A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms* **6**, 161–179 (1995)
- [26] Newman, M.E.J.: Assortative mixing in networks. *Physical Review Letters* **89**(20), 208,701 (2002)
- [27] Newman, M.E.J.: Spread of epidemic disease on networks. *Physical Review E* **66**(1), 016,128 (2002)
- [28] O’Neill, P.D., Roberts, G.O.: Bayesian inference for partially observed stochastic epidemics. *Journal of the Royal Statistical Society A* **162**, 121–129 (1999)
- [29] Parsons, T.L., Rogers, T.: Dimension reduction via timescale separation in stochastic dynamical systems (2015). [arXiv:1510.07031]
- [30] Ross, J.V., Taimre, T., Pollett, P.K.: On parameter estimation in population models. *Theoretical Population Biology* **70**(4), 498–510 (2006)
- [31] Volz, E.M.: SIR dynamics in random networks with heterogeneous connectivity. *Journal of Mathematical Biology* **56**(3), 293–310 (2008)
- [32] Waugh, W.A.O.: Conditioned Markov processes. *Biometrika* **45**(1-2), 241–249 (1958)
- [33] Windridge, P.: The extinction time of a subcritical branching process related to the SIR epidemic on a random graph (2014). To appear in *Journal of Applied Probability*. [arXiv:1408.4768]

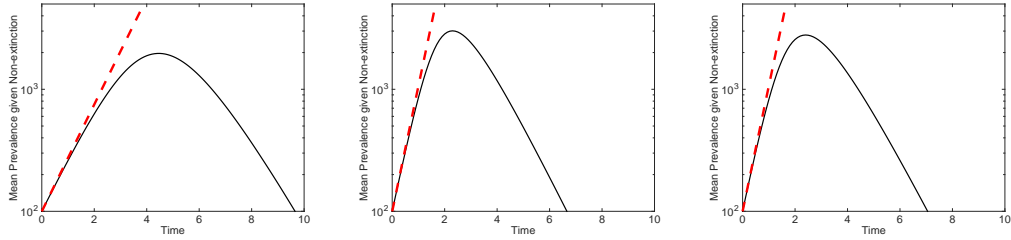
(i) Degree distribution histograms.



(ii) 100 sample trajectories.



(iii) Mean prevalence. Black solid: simulations; Red dashed: branching process.



(iv) Variance in prevalence. Black solid: simulations; Red dashed: branching process.

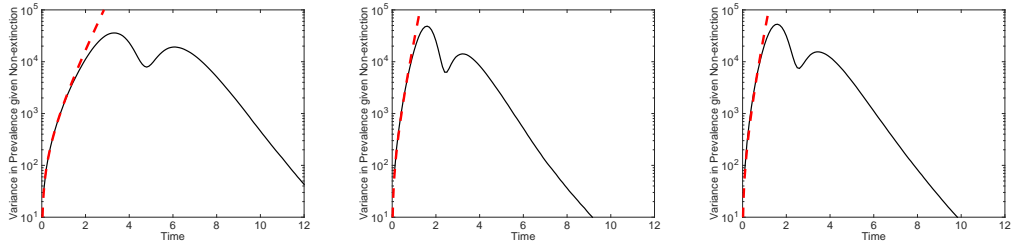
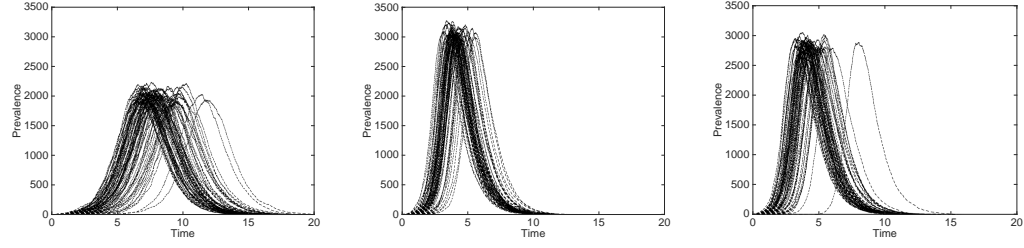
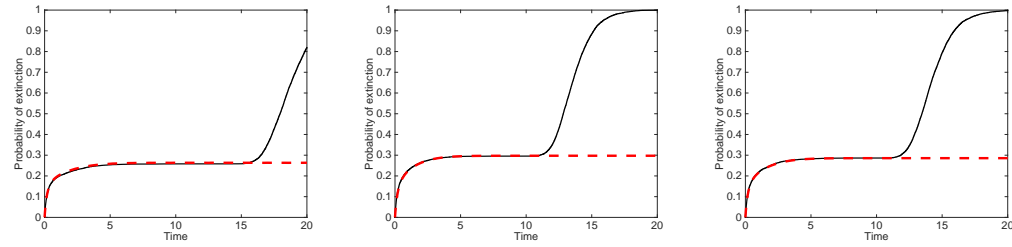


Figure 1: Epidemic simulations that set time = 0 when prevalence is equal to 100. Parameters are $\tau = 2$, $\gamma = 1$ throughout.

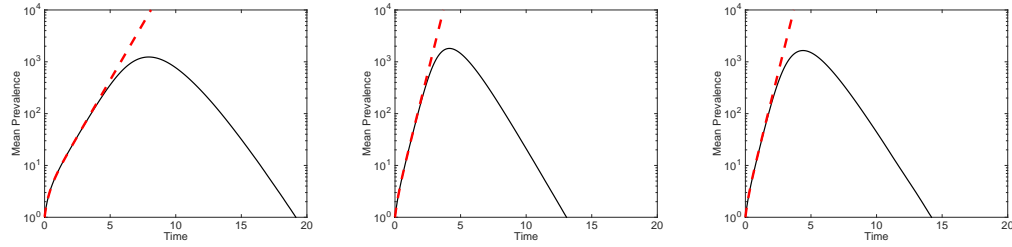
(i) 100 sample trajectories.



(ii) Extinction probabilities. Black solid: simulations; Red dashed: branching process.



(iii) Mean prevalence. Black solid: simulations; Red dashed: branching process.



(iv) Variance in prevalence. Black solid: simulations; Red dashed: branching process.

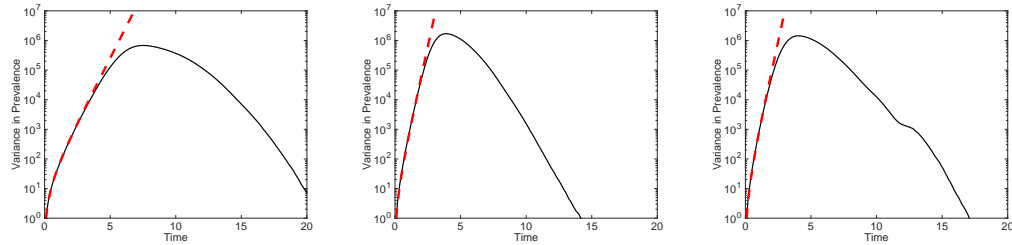
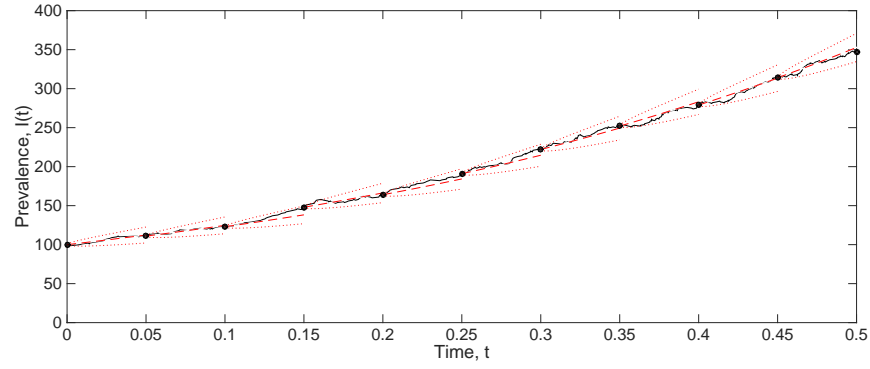


Figure 2: Epidemic simulations starting from one node selected uniformly at random. Parameters are $\tau = 2$, $\gamma = 1$ throughout. Degree distributions are as for Figure 1 above.

(i) Simulated epidemic



(ii) Likelihood surfaces

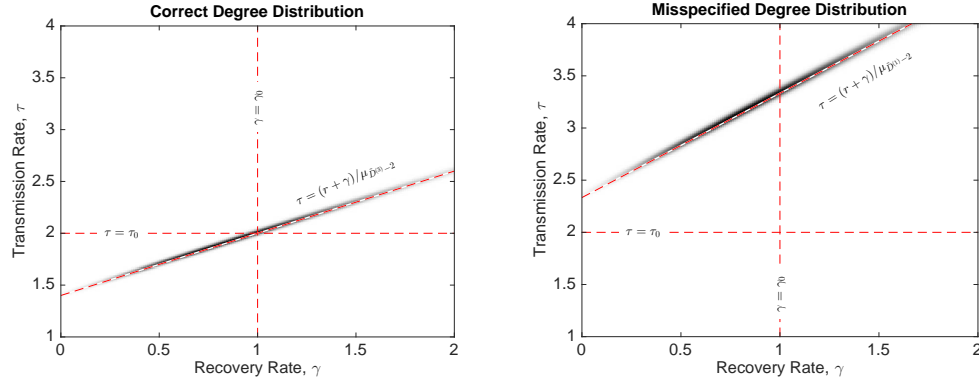


Figure 3: Simulation study. The top plot (i) shows the first quarter of the timepoints (observations as black dots, full trajectory as black solid line, Gaussian approximation mean as red dashed line, Gaussian approximation 95% prediction interval as red dotted line). The bottom plots (ii) show likelihood surfaces for a Gaussian approximation model as described in Section 9.2. True parameters are $\hat{\tau} = 2$, $\hat{\gamma} = 1$. Correct degree distribution $D^{(3)}$ is as given in the third columns of Figures 1 and 2 above, and the misspecified distribution $D^{(1)}$ is the distribution from the first column. Likelihood at a point is proportional to the intensity of shading, and three curves are shown in each plot as dashed red lines.